

GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.  
Source version 3.1.6

OM protein - protein search, using sw model

Run on: July 7, 2003, 14:50:02 ; Search time 70 Seconds  
(without alignments)

Title: US-09-591-500A-4  
 perfect score: 1216  
 sequence: 1 MEMGRIHSELNRPADVK.....VDGEDEELGEERQGRK 234

Scoring table: BLOSUM62

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45

```

database :
1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query % |       | Length | DB       | ID                 | Description |
|------------|-------|---------|-------|--------|----------|--------------------|-------------|
|            |       | Match   | Match |        |          |                    |             |
| 1          | 1216  | 100.0   | 234   | 20     | AAV16623 | Human phosphoprote |             |
| 2          | 1216  | 100.0   | 234   | 20     | AAV22800 | Amino acid sequenc |             |
| 3          | 1216  | 100.0   | 234   | 21     | AAE20649 | Human pp32 protein |             |
| 4          | 1216  | 100.0   | 234   | 21     | AAE20655 | Human variant pp32 |             |
| 5          | 1209  | 99.4    | 234   | 20     | AAV16622 | Human phosphoprote |             |
| 6          | 1209  | 99.4    | 234   | 21     | AAE20648 | Human pp32 protein |             |
| 7          | 1161  | 95.5    | 245   | 20     | AAV16624 | Human phosphoprote |             |
| 8          | 1161  | 95.5    | 245   | 21     | AAE20650 | Human pp32 protein |             |
| 9          | 1001  | 82.3    | 249   | 17     | AAE95900 | Human pp32 protein |             |
| 10         | 1001  | 82.3    | 249   | 20     | AAV16626 | Human phosphoprote |             |

## ALIGNMENTS

## RESULT 1

AAAY16623

ID AAY16623 standard; Protein; 234 AA.

AA  
AC  
AAY16623:

XX

26-AUG-1999 (first entry)

100

human phosphoprotein 32 (pp32) protein sequence.

Phosphoprotein 32; pp32; variant; pp32r1; pp32r2; cancer; prostate; prostatic adenocarcinoma; antineoplastic activity; PCR primer; transformation suppression; malignant potential; neuroendocrine; neural; mesenchymal; lymphoid; epithelial; germ cell; tumour.

Homo sapiens.

XX

WO9929906-A2.

17-TIN-1000

17-JUN-1999.

11-DEC-1998; 98WO-US26433.

XIX

12-DEC-1997; 9705-0069677.

PA (UYJO ) UNIV JOHNS HOPKINS.

X  
I

И-И-И, Кадко! СС

R WPI; 1999-385626/32.

X E

1 X F103 protein 32 (pp32) related genomic sequences

PS Example 6; Page 46-47; 65pp; English.

XX AAY16615-28 represent human phosphoprotein 32 (pp32) protein sequences  
CC isolated from different patients. The specification describes pp32  
CC variants, designated pp32r1 and pp32r2. The pp32r1 and pp32r2 sequences  
CC are associated with cancer in prostate, especially prostatic  
CC adenocarcinomas. Normal pp32 exerts antineoplastic activity through  
CC suppression of transformation. Cancer-associated pp32 variants augment,  
CC rather than inhibit, transformation. Determining the presence of a gene  
CC encoding residues 146-163 of pp32r1 or pp32r2 in a sample is useful for  
CC a diagnostic method for predicting malignant potential of  
CC neuroendocrine, neural, mesenchymal, lymphoid, epithelial or germ  
CC cell-derived tumours.

XX SQ Sequence 234 AA;  
Query Match 100.0%; Score 1216; DB 20; Length 234;  
Best Local Similarity 100.0%; Pred. No. 3.5e-98;  
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNEKLEALTDFFEFLSKINGLTSISDL 60  
DB 1 MEMGRIHSELNRAPSDVKELALDNRSGNEKLEALTDFFEFLSKINGLTSISDL 60  
QY 61 PKLKRLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKSLDLFNC 120  
DB 61 PKLKRLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKSLDLFNC 120  
QY 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDDEEGEHEEYDEDAQ 180  
DB 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDDEEGEHEEYDEDAQ 180  
QY 181 VVEDEGEEREEEGEEDVSGGDEDEEGYNDGEVDEDEELGEEERQKRK 234  
DB 181 VVEDEGEEREEEGEEDVSGGDEDEEGYNDGEVDEDEELGEEERQKRK 234

## RESULT 2

AAY22800  
ID AAY22800 standard; Protein; 234 AA.

XX AC AAY22800;

XX DT 26-AUG-1999 (first entry)

XX DE Amino acid sequence of phosphoprotein 32 variant pp32r1.

XX KW Phosphoprotein 32; pp32; variant; pp32r1; pp32r2; cancer; prostate;  
XX KW prostatic adenocarcinoma; antineoplastic activity;  
XX KW transformation suppression; malignant potential; neuroendocrine;  
XX KW neural; mesenchymal; lymphoid; epithelial; germ cell; tumour.

XX OS Homo sapiens.

XX XX WO9929906-A2.

XX PN 17-JUN-1999.

XX XX 11-DEC-1998; 98WO-US26433.

XX XX 12-DEC-1997; 97US-0069677.

XX XX (UYJO ) UNIV JOHNS HOPKINS.

XX XX Brody JR, Kadkol SS, Kocheavar GJ, Pasternack GR;

XX XX WPI; 1999-385626/32.

XX XX Phosphoprotein 32 (pp32) related genomic sequences

XX XX Example 2; Fig 4; 65pp; English.

XX XX The present sequence represents the phosphoprotein 32 (pp32)

CC variant pp32r1, which is isolated from human placenta. The  
CC pp32r1 and pp32r2 sequences are associated with cancer in prostate,  
CC especially prostatic adenocarcinomas. Normal pp32 exerts antineoplastic  
CC activity through suppression of transformation. Cancer-associated pp32  
CC variants augment, rather than inhibit, transformation. Determining the  
CC presence of a gene encoding residues 146-163 of pp32r1 or pp32r2 in a  
CC sample is useful for a diagnostic method for predicting malignant  
CC potential of neuroendocrine, neural, mesenchymal, lymphoid, epithelial  
CC or germ cell-derived tumours.

XX SQ Sequence 234 AA;  
Query Match 100.0%; Score 1216; DB 20; Length 234;  
Best Local Similarity 100.0%; Pred. No. 3.5e-98;  
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNEKLEALTDFFEFLSKINGLTSISDL 60  
DB 1 MEMGRIHSELNRAPSDVKELALDNRSGNEKLEALTDFFEFLSKINGLTSISDL 60  
QY 61 PKLKRLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKSLDLFNC 120  
DB 61 PKLKRLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKSLDLFNC 120  
QY 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDDEEGEHEEYDEDAQ 180  
DB 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDDEEGEHEEYDEDAQ 180  
QY 181 VVEDEGEEREEEGEEDVSGGDEDEEGYNDGEVDEDEELGEEERQKRK 234  
DB 181 VVEDEGEEREEEGEEDVSGGDEDEEGYNDGEVDEDEELGEEERQKRK 234

## RESULT 3

AAB20649  
ID AAB20649 standard; Protein; 234 AA.

XX AC AAB20649;

XX DT 15-DEC-2000 (first entry)

XX DE Human pp32 protein sequence clone KG.

XX KW Human; pp32; chromosome 15q22.3-q23; cancer; phosphoprotein; APRIL;  
XX KW acidic protein rich in leucine; chromosome 15q25; tumour suppressor;  
XX KW malignant; cytostatic; gene therapy.

XX OS Homo sapiens.

XX XX WO200045852-A1.

XX XX 10-AUG-2000.

XX XX 03-FEB-2000; 2000WO-US02656.

XX XX 03-FEB-1999; 99US-0118667.

XX XX (UYJO ) UNIV JOHNS HOPKINS.

XX XX Pasternack GR, Bai J;

XX XX WPI; 2000-514896/46.

XX XX Treatment of cancer comprising restoration of pp32 function in  
XX PT malignant cells -

XX PS Example 6; Page 49-50; 90pp; English.

XX CC The present invention describes a method (M1) for treating malignant  
XX CC cells comprising restoration of pp32 function. Also described are:  
XX CC (1) a method (M2) of screening to determine whether a compound is an  
XX CC inducer of pp32 expression comprising measuring pp32 expression by  
XX CC cells cultured in the presence and absence of the compound; and

CC (2) a method (M3) of screening to determine whether a compound is an  
CC inducer of pp32 function comprising measuring protein phosphatase  
CC activity in cells cultured in the presence and absence of the compound.  
CC The methods are useful for treating cancer and for identifying agents  
CC which may be used to treat cancer. Human pp32 is a phosphoprotein which  
CC has been mapped to chromosome 15q22.3-q23. The present sequence  
CC represents a human pp32 protein sequence from an example from the  
CC present invention.  
XX  
SQ Sequence 234 AA;

Query Match 100.0%; Score 1216; DB 21; Length 234;  
Best Local Similarity 100.0%; Pred. No. 3.5e-98;  
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNEGKLEALTDFEELFSLKINGLTSIDL 60  
Db 1 MEMGRIHSELNRAPSDVKELALDNRSGNEGKLEALTDFEELFSLKINGLTSIDL 60  
QY 61 PKLKRKLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENLSLDLNFCE 120  
Db 61 PKLKRKLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENLSLDLNFCE 120  
QY 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
Db 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
QY 181 VVEDEGEHEEHEEEDVSGDDEDEGYNDGEVDEDEEELGEEERQKRK 234  
Db 181 VVEDEGEHEEHEEEDVSGDDEDEGYNDGEVDEDEEELGEEERQKRK 234

RESULT 4  
AAB20655  
ID AAB20655 standard; Protein; 234 AA.

XX AAB20655;  
AC AAB20655;

DT 15-DEC-2000 (first entry)

DE Human variant pp32r1 protein sequence.

XX Human; pp32; chromosome 15q22.3-q23; cancer; phosphoprotein; APRIL;  
KW acidic protein rich in leucine; Chromosome 15q25; tumour suppressor;  
KW malignant; cytostatic; gene therapy.

XX Homo sapiens.

OS WO200045852-A1.

PN 10-AUG-2000.

PD 03-FEB-2000; 2000WO-US02656.

PF 03-FEB-1999; 99US-0118667.

PR (UYJO ) UNIV JOHNS HOPKINS.

PA Pasternack GR, Bai J;

PI WPI; 2000-514896/46.

DR N-PSDB; AAA88237.

XX Treatment of cancer comprising restoration of pp32 function in  
PT malignant cells -

XX Example 3; Fig 4; 90pp; English.

XX The present invention describes a method (M1) for treating malignant  
CC cells comprising restoration of pp32 function. Also described are:  
CC (1) a method (M2) of screening to determine whether a compound is an  
CC inducer of pp32 expression comprising measuring pp32 expression by  
CC cells cultured in the presence and absence of the compound; and

CC (2) a method (M3) of screening to determine whether a compound is an  
CC inducer of pp32 function comprising measuring protein phosphatase  
CC activity in cells cultured in the presence and absence of the compound.  
CC The methods are useful for treating cancer and for identifying agents  
CC which may be used to treat cancer. Human pp32 is a phosphoprotein which  
CC has been mapped to chromosome 15q22.3-q23. The present sequence  
CC represents the human variant pp32r1 protein sequence from  
CC an example of the present invention.  
XX  
SQ Sequence 234 AA;

Query Match 100.0%; Score 1216; DB 21; Length 234;  
Best Local Similarity 100.0%; Pred. No. 3.5e-98;  
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNEGKLEALTDFEELFSLKINGLTSIDL 60  
Db 1 MEMGRIHSELNRAPSDVKELALDNRSGNEGKLEALTDFEELFSLKINGLTSIDL 60  
QY 61 PKLKRKLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENLSLDLNFCE 120  
Db 61 PKLKRKLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENLSLDLNFCE 120  
QY 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
Db 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
QY 181 VVEDEGEHEEHEEEDVSGDDEDEGYNDGEVDEDEEELGEEERQKRK 234  
Db 181 VVEDEGEHEEHEEEDVSGDDEDEGYNDGEVDEDEEELGEEERQKRK 234

RESULT 5

AAVI6622  
ID AAVI6622 standard; Protein; 234 AA.

XX AAVI6622;  
AC AAVI6622;

DT 26-AUG-1999 (first entry)

DE Human phosphoprotein 32 (pp32) protein sequence.

XX Phosphoprotein 32; pp32; variant; pp32r1; pp32r2; cancer; prostate;  
KW prostatic adenocarcinoma; antineoplastic activity; PCR primer;  
KW transformation suppression; malignant potential; neuroendocrine;  
KW neural; mesenchymal; lymphoid; epithelial; germ cell; tumour.

XX Homo sapiens.

OS WO929906-A2.

PN 17-JUN-1999.

PD 11-DEC-1998; 98WO-US26433.

PF 12-DEC-1997; 97US-0069677.

PR (UYJO ) UNIV JOHNS HOPKINS.

PA Brody JR, Kadkol SS, Kochevar GJ, Pasternack GR;

PI WPI; 1999-385626/32.

DR Phosphoprotein 32 (pp32) related genomic sequences

XX Example 6; Page 46-47; 65pp; English.

XX AAVI6615-28 represent human phosphoprotein 32 (pp32) protein sequences  
CC isolated from different patients. The specification describes pp32  
CC variants, designated pp32r1 and pp32r2. The pp32r1 and pp32r2 sequences  
CC are associated with cancer in prostate, especially prostatic  
CC adenocarcinomas. Normal pp32 exerts antineoplastic activity through  
CC suppression of transformation. Cancer-associated pp32 variants augment,

CC rather than inhibit, transformation. Determining the presence of a gene  
CC encoding residues 146-163 of pp32r1 or pp32r2 in a sample is useful for  
CC a diagnostic method for predicting malignant potential of  
CC neuroendocrine, neural, mesenchymal, lymphoid, epithelial or germ  
CC cell-derived tumours.

XX SQ Sequence 234 AA;  
Query Match 99.4%; Score 1209; DB 20; Length 234;  
Best Local Similarity 99.1%; Pred. No. 1.4e-97;  
Matches 232; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNEKLEALTDDEFELEFLSKINGLTSIDL 60  
DB 1 MEMGRIHSELNRAPSDVKELVLDNRSGNEKLEALTDDEFELEFLSKINGLTSIDL 60  
QY 61 PKLKLRLKLELVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKOLENLSKLDLFNCE 120  
DB 61 PKLKLRLKLELVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKOLENLSKLDLFNCE 120  
QY 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
DB 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
QY 181 VVEDEGEHEEYDEEYSGGDEDEEGYNDGVDGEDEEELGEEERQKRR 234  
DB 181 VVEDEGEHEEYDEEYSGGDEDEEGYNDGVDGEDEEELGEEERQKRR 234

RESULT 6  
AAB20648  
ID AAB20648 standard; Protein; 234 AA.  
XX AC AAB20648;  
XX DT 15-DEC-2000 (first entry)  
XX DE Human pp32 protein sequence clone FT2.2.  
XX KW Human; pp32; chromosome 15q22.3-q23; cancer; phosphoprotein; APRIL;  
KW acidic protein rich in leucine; chromosome 15q25; tumour suppressor;  
KW malignant; cytostatic; gene therapy.  
XX OS Homo sapiens.  
XX PN WO200045852-A1.  
XX PD 10-AUG-2000.  
XX PF 03-FEB-2000; 2000WO-US02656.  
XX PR 03-FEB-1999; 99US-0118667.  
XX PA (UYJO ) UNIV JOHNS HOPKINS.  
XX PI Pasternack GR, Bai J;  
XX DR WPI; 2000-514896/46.  
XX PT Treatment of cancer comprising restoration of pp32 function in  
XX malignant cells -  
XX PS Example 6; Page 49-50; 90pp; English.

XX The present invention describes a method (M1) for treating malignant  
XX cells comprising restoration of pp32 function. Also described are:  
XX (1) a method (M2) of screening to determine whether a compound is an  
XX inducer of pp32 expression comprising measuring pp32 expression by  
XX cells cultured in the presence and absence of the compound; and  
XX (2) a method (M3) of screening to determine whether a compound is an  
XX inducer of pp32 function comprising measuring protein phosphatase  
XX activity in cells cultured in the presence and absence of the compound.  
XX The methods are useful for treating cancer and for identifying agents

CC which may be used to treat cancer. Human pp32 is a phosphoprotein which  
CC has been mapped to chromosome 15q22.3-q23. The present sequence  
CC represents a human pp32 protein sequence from an example from the  
CC present invention.

XX SQ Sequence 234 AA;  
Query Match 99.4%; Score 1209; DB 21; Length 234;  
Best Local Similarity 99.1%; Pred. No. 1.4e-97;  
Matches 232; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNEKLEALTDDEFELEFLSKINGLTSIDL 60  
DB 1 MEMGRIHSELNRAPSDVKELVLDNRSGNEKLEALTDDEFELEFLSKINGLTSIDL 60  
QY 61 PKLKLRLKLELVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKOLENLSKLDLFNCE 120  
DB 61 PKLKLRLKLELVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKOLENLSKLDLFNCE 120  
QY 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
DB 121 VTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
QY 181 VVEDEGEHEEYDEEYSGGDEDEEGYNDGVDGEDEEELGEEERQKRR 234  
DB 181 VVEDEGEHEEYDEEYSGGDEDEEGYNDGVDGEDEEELGEEERQKRR 234

RESULT 7  
AA116624  
ID AAY16624 standard; Protein; 245 AA.  
XX AC AAY16624;  
XX DT 26-AUG-1999 (first entry)  
XX DE Human phosphoprotein 32 (pp32) protein sequence.  
XX KW Phosphoprotein 32; pp32; variant; pp32r1; pp32r2; cancer; prostate;  
KW prostatic adenocarcinoma; antineoplastic activity; PCR primer;  
KW transformation suppression; malignant potential; neuroendocrine;  
KW neural; mesenchymal; lymphoid; epithelial; germ cell; tumour.  
XX OS Homo sapiens.  
XX PN WO9929906-A2.  
XX PD 17-JUN-1999.  
XX PF 11-DEC-1998; 98WO-US26433.  
XX PR 12-DEC-1997; 97US-0069677.  
XX PA (UYJO ) UNIV JOHNS HOPKINS.  
XX PI Brody JR, Kadkol SS, Kocheavar GJ, Pasternack GR;  
XX DR WPI; 1999-385626/32.  
XX PT Phosphoprotein 32 (pp32) related genomic sequences  
XX PS Example 6; Page 46-47; 65pp; English.

XX AAY16615-28 represent human phosphoprotein 32 (pp32) protein sequences  
XX isolated from different patients. The specification describes pp32  
XX variants, designated pp32r1 and pp32r2. The pp32r1 and pp32r2 sequences  
XX are associated with cancer in prostate, especially prostatic  
XX adenocarcinomas. Normal pp32 exerts antineoplastic activity through  
XX suppression of transformation. Cancer-associated pp32 variants augment,  
XX rather than inhibit, transformation. Determining the presence of a gene  
XX encoding residues 146-163 of pp32r1 or pp32r2 in a sample is useful for  
XX a diagnostic method for predicting malignant potential of  
XX neuroendocrine, neural, mesenchymal, lymphoid, epithelial or germ

CC cell-derived tumours.

XX SQ Sequence 245 AA;

Query Match 95.5%; Score 1161; DB 20; Length 245;  
Best Local Similarity 95.7%; Pred. No. 2.4e-93;  
Matches 224; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 MEMGRRTHSLRNAPSVDVKELALDNRSGNEKLEALTDPEFELEPLSKINGLTSIDL 60  
DB 1 MEMGRRHLELRNRTPSDVKELVLDNRSGNEKLEGLTDEFEFELEFLSTINVGLTSIANL 60  
QY 61 PKLKRKLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKQLENKSLDLFNCE 120  
DB 61 PKLKRKLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKQLENKSLDLFNCE 120  
QY 121 VTNLNDYGENVFKLLQLTYLDSYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
DB 121 VTNLNDYGENVFKLLQLTYLDSYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
QY 181 VVEDEGEGEREEDVSGGDEDEEGYNDGEVDEDEELGEEERGGQRK 234  
DB 181 VVEDEGEGEREEDVSGGDEDEEGYNDGEVDEDEELGEEERGGQRK 234

RESULT 8

AAB20650

ID AAB20650 standard; Protein; 245 AA.

XX AC AAB20650;

XX DT 15-DEC-2000 (first entry)

XX DE Human pp32 protein sequence clone FT1.7.

XX KW Human; pp32; chromosome 15q22.3-q23; cancer; phosphoprotein; APRIL;  
XX KW acidic protein rich in leucine; chromosome 15q25; tumour suppressor;  
XX KW malignant; cytostatic; gene therapy.

XX OS Homo sapiens.

XX PN W0200045852-A1.

XX PD 10-AUG-2000.

XX PF 03-FEB-2000; 2000WO-US02656.

XX PR 03-FEB-1999; 99US-0118667.

XX PA (UYJO ) UNIV JOHNS HOPKINS.

XX PI Pasternack GR, Bai J;

XX DR WPI; 2000-514896/46.

XX PT Treatment of cancer comprising restoration of pp32 function in  
XX PT malignant cells -

XX PS Example 6; Page 49-50; 90pp; English.

XX CC The present invention describes a method (M1) for treating malignant  
XX CC cells comprising restoration of pp32 function. Also described are:  
XX CC (1) a method (M2) of screening to determine whether a compound is an  
XX CC inducer of pp32 expression comprising measuring pp32 expression by  
XX CC cells cultured in the presence and absence of the compound; and  
XX CC (2) a method (M3) of screening to determine whether a compound is an  
XX CC inducer of pp32 function comprising measuring protein phosphatase  
XX CC activity in cells cultured in the presence and absence of the compound.  
XX CC The methods are useful for treating cancer and for identifying agents  
XX CC which may be used to treat cancer. Human pp32 is a phosphoprotein which  
XX CC has been mapped to chromosome 15q22.3-q23. The present sequence  
XX CC represents a human pp32 protein sequence from an example from the  
XX CC present invention.

XX SQ Sequence 245 AA;

Query Match 95.5%; Score 1161; DB 21; Length 245;  
Best Local Similarity 95.7%; Pred. No. 2.4e-93;  
Matches 224; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 MEMGRRHSELRNAPSVDVKELALDNRSGNEKLEALTDPEFELEFLSKINGLTSIDL 60  
DB 1 MEMGRRHLELRNRTPSDVKELVLDNRSGNEKLEGLTDEFEFELEFLSTINVGLTSIANL 60  
QY 61 PKLKRKLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKQLENKSLDLFNCE 120  
DB 61 PKLKRKLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKQLENKSLDLFNCE 120  
QY 121 VTNLNDYGENVFKLLQLTYLDSYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
DB 121 VTNLNDYGENVFKLLQLTYLDSYWDHKEAPYSDIEDHVEGLDDEEGEHEEYDEDAQ 180  
QY 181 VVEDEGEGEREEDVSGGDEDEEGYNDGEVDEDEELGEEERGGQRK 234  
DB 181 VVEDEGEGEREEDVSGGDEDEEGYNDGEVDEDEELGEEERGGQRK 234

RESULT 9

AAR95900

ID AAR95900 standard; Protein; 249 AA.

XX AC AAR95900;

XX DT 30-JUL-1996. (first entry)

XX DE Human pp32 protein.

XX KW pp32; cancer; diagnosis; therapy; antisense; cell proliferation;  
XX KW lymphoid tumour; epithelial tumour; colon carcinoma;  
XX KW prostate carcinoma; non-Hodgkin lymphoma; apoptosis.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FD Domain 1..167

XX FT /note= "N-terminal domain is generally amphipathic  
XX FT and has a high probability of alpha-helix  
XX FT conformation"

XX FT Region 69..90

XX FT /note= "candidate leucine zipper composed of  
XX FT Leu-69, Val-76, Leu-83 and Leu-90"

XX FT Region 60..68

XX FT /note= "candidate nuclear localisation sequence"  
XX FT 108..116

XX FT /note= "candidate nuclear localisation sequence"  
XX FT 168..249

XX FT /note= "highly acidic C-terminal domain"

XX PN W09610092-A1.

XX PD 04-APR-1996.

XX PF 28-SEP-1995; 95WO-US12414.

XX PR 28-SEP-1994; 94US-0314503.

XX PA (UYJO ) UNIV JOHNS HOPKINS.

XX PI Kuhajda FP, Pasternack GR;

XX DR WPI; 1996-200930/20.

XX DR N-PSDB; AAT27712.

XX PT New method of diagnosing cancer using pp32 cDNA - by detecting the  
XX PT level of mRNA hybridising to pp32 cDNA; also for inhibiting cell  
XX PT proliferation and screening anti-cancer drugs.





PT New isolated V3 loop HIV receptor - comprises P95/nucleolin,  
PT P40/PHAPII and P30/PHAPI proteins, used to develop products for the  
PT treatment and prevention of HIV infection  
XX  
PS Disclosure: Fiq 49(11); 267pp; English.

| Query Match           | 82.38;          | Score 1001;  | DB 20;    | Length 249; |
|-----------------------|-----------------|--|-----------|-------------|
| Best Local Similarity | 86.18;          | Pred. No. 2.2e-79;   |           |             |
| Matches 205;          | Conservative 7; | Mismatches 22;   | Indels 4; | Gaps 2;     |
| QY                    | 1               | MEMGRIHSELNRAPSDVKELALDNSRSNEGKLEALDTEFEELFLSKINGLGLTSIDL  | 60        |             |
| DB                    | 1               | MEMGRIHLELRNTPSDVKELVLDNSRSNEGKLEGLTDEFEELFLSTINVLGLTSIANL | 60        |             |
| QY                    | 61              | PKL-KLRKLEI---RVSGGLEVLAEKCPNTHLYLSGNKIKDLSTIEPKULENKLSIDL | 116       |             |
| DB                    | 61              | PKLNKLRKLEISDRNVSGGLEVLAEKCPNTHLNLSGNKIKDLSTIEPKULENKLSIDL | 120       |             |
| QY                    | 117             | FNCEVTNLNDXGNVFKLLQLLTALDYDSCYWDHKEAPYSDIEDHVEGLDDEEHEEYD  | 176       |             |
| DB                    | 121             | FNCEVTNLNDYRNVFKLLQLTYLDGYDRDKEAPSDAEGYVGLDDEEDEDYD        | 180       |             |
| QY                    | 177             | EDAQVVEDEGEHEEEDYSGGDEDEEGYNDGVGDDEBELGEERGGQRK            | 234       |             |
| DB                    | 181             | EDAQVVEDEDEDEGEEDYSGGDEDEEGYNDGVGDDEBELGEERGGQRK           | 238       |             |

|           |  |
|-----------|--|
| RESULT 14 |  |
| AAB20652  |  |
| ID        | AAB20652 standard; Protein; 249 AA.                                  |
| XX        |  |
| XX        | AAB20652;  |
| XX        |  |
| XX        | 15-DEC-2000 (first entry)  |
| DT        |  |
| DT        |  |
| XX        | Human pp32 protein sequence clone L3.                                |
| DE        |  |
| XX        |  |
| XX        | Human; pp32; chromosome 15q22.3-q23; cancer; phosphoprotein; APRIL;  |
| KW        | acidic protein rich in leucine; Chromosome 15q25; tumour suppressor; |
| KW        | malignant; cytostatic; gene therapy.                                 |
| XX        |  |
| XX        |  |
| OS        | Homo sapiens.  |
| XX        |  |
| PN        | WO200045852-A1.  |
| XX        |  |
| PD        | 10-AUG-2000.   |
| XX        |  |
| PF        | 03-FEB-2000; 2000WO-US02656.   |
| XX        |  |
| PR        | 03-FEB-1999; 99US-0118667.   |
| PR        |  |

XX (UYJO ) UNIV JOHNS HOPKINS.  
PA Pasternack GR, Bai J;  
XX WPI; 2000-514896/46.  
XX Treatment of cancer comprising restoration of pp32 function in  
PT malignant cells -  
XX Example 6; Page 49-50; 90pp; English.  
PS  
XX The present invention describes a method (M1) for treating malignant  
CC cells comprising restoration of pp32 function. Also described are:  
CC (1) a method (M2) of screening to determine whether a compound is an  
CC inducer of pp32 expression comprising measuring pp32 expression by  
CC cells cultured in the presence and absence of the compound; and  
CC (2) a method (M3) of screening to determine whether a compound is an  
CC inducer of pp32 function comprising measuring protein phosphatase  
CC activity in cells cultured in the presence and absence of the compound.  
CC The methods are useful for treating cancer and for identifying agents  
CC which may be used to treat cancer. Human pp32 is a phosphoprotein which  
CC has been mapped to chromosome 15q22.3-q23. The present sequence  
CC represents a human pp32 protein sequence from an example from the  
CC present invention.  
XX  
XX Sequence 249 AA;  
SQ

|    | Query Match           | 82.38;   | Score 1001;        | DB 21;         | Length 249;       |
|----|-----------------------|--|--------------------|----------------|-------------------|
|    | Best Local Similarity | 86.1%;   | Pred. No. 2.2e-79; |                |                   |
|    | Matches 205;          | Conservative   | 7;                 | Mismatches 22; | Indels 4; Gaps 2; |
| Qy | 1                     | MEMGRRTHSELNRAPSDVKELALDNRSGNEGKLEALTDFEELEFLSKINGGLTSSIDL   | 60                 |                |                   |
| Db | 1                     | MEMGRRTHLELRNRTPSDVKELVLDNRSGNEGKLEGLTDFEELEFLSTINVGLTSTIANL | 60                 |                |                   |
| Qy | 61                    | PKL-KLKKLEL---RVSGGLEVLAEKCPNTHLYLSGNKIKDLSIETPLQLKNLKSIDL   | 116                |                |                   |
| Db | 61                    | PKLKLKKLELSDNRVSGGLEVLAEKCPNTHNLNSGNKIKDLSIETPLKKLENLKSIDL   | 120                |                |                   |
| Qy | 117                   | FNCEVTNLNDYGENVFLLQLTLYLDCYWDHKAPYSDIEDHVEGLDDDEEGHSEYD      | 176                |                |                   |
| Db | 121                   | FNCEVTNLNDYRENVFKLLPQLTLYLDGYDRDKAPDSAGYVEGLDDDEEDEYD        | 180                |                |                   |
| Qy | 177                   | EDAQVDEDEGEFEESGEEDVSGGDEDEEGYNGDGEDEDEELGGEERQKKR           | 234                |                |                   |
| Db | 181                   | EDAQWDEDEDEFEESGEEDVSGGEDEEGYNGDGEDEDEDEELGGEERQKKR          | 238                |                |                   |

|           |   |
|-----------|---|
| RESULT 15 |   |
| AAB20653  |   |
| ID        | AAB20653 standard; Protein; 249 AA.     |
| XX        |   |
| XX        |   |
| AC        | AAB20653;                               |
| XX        |   |
| DT        | 15-DEC-2000 (first entry)               |
| XX        |   |
| DE        | Human pp32 protein sequence clone pp32. |
| XX        |   |
| XX        |   |
| KW        | Human; pp32; chromosome 15q22.3-q23; ca |
| KW        | acidic protein rich in leucine; chromos |
| KW        | malnigant; cytostatic; gene therapy.    |
| XX        |   |
| OS        | Homo sapiens.                           |
| XX        |   |
| PN        | WO200045852-A1.                         |
| XX        |   |
| PD        | 10-AUG-2000.                            |
| XX        |   |
| PF        | 03-FEB-2000; 2000WO-US02656.            |
| XX        |   |
| PR        | 03-FEB-1999; 99US-0118667.              |
| XX        |   |









```
Db 1 MEMGRIHLELRNRTSDVKELVLDNSRNEKLEGLTDFEFLSTINVGLTSIDL 60
Qy 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENKSIDL 116
Db 61 PKLKLKLELSDNRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENKSIDL 120
Qy 117 FNCVTLNDYGNVFKLLQLTLYLSDCYWDHKEAPYSDIEHVEGLDDEEGEHEEYD 176
Db 121 FNCVTLNDYGNVFKLLQLTLYLSDCYWDHKEAPYSDIEHVEGLDDEEGEHEEYD 180
Qy 177 EDAQVVEDEEEDVEEEDVSGGDEDEEGYNDGVDGDEDEEELGEEERQKRK 234
Db 181 EDAQVVEDEEEDVEEEDVSGGDEDEEGYNDGVDGDEDEEELGEEERQKRK 238

RESULT 2
US-08-314-503A-2
; Sequence 2, Application US/08314503A
; Patent No. 5734022
; GENERAL INFORMATION:
; APPLICANT: Pasternack, Gary R.
; APPLICANT: Kuhaajda, Francis P.
; TITLE OF INVENTION: No. 5734022el Mammalian Protein Associated With
; TITLE OF INVENTION: Uncontrolled Cell Division
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington, D.C.
; STATE: District of Columbia
; COUNTRY: U.S.A.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/314,503A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Posorske Esq., Laurence H.
; REGISTRATION NUMBER: 34,698
; REFERENCE/DOCKET NUMBER: 1107.47218
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 508-9153
; TELEFAX: 202 508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-314-503A-2

Query Match 82.3%; Score 1001; DB 1; Length 249;
Best Local Similarity 86.1%; Pred. No. 3.5e-88;
Matches 205; Conservative 7; Mismatches 22; Indels 4; Gaps 2;

Qy 1 MEMGRIHSELRNAPSVDKELALDNSRNEKLEALTDFEFLSKINGLTSIDL 60
Db 1 MEMGRIHLELRNRTSDVKELVLDNSRNEKLEGLTDFEFLSTINVGLTSIDL 60
Qy 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENKSIDL 116
Db 61 PKLKLKLELSDNRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENKSIDL 120
Qy 117 FNCVTLNDYGNVFKLLQLTLYLSDCYWDHKEAPYSDIEHVEGLDDEEGEHEEYD 176
Db 121 FNCVTLNDYGNVFKLLQLTLYLSDCYWDHKEAPYSDIEHVEGLDDEEGEHEEYD 180
Qy 177 EDAQVVEDEEEDVEEEDVSGGDEDEEGYNDGVDGDEDEEELGEEERQKRK 234
Db 181 EDAQVVEDEEEDVEEEDVSGGDEDEEGYNDGVDGDEDEEELGEEERQKRK 238

RESULT 4
US-08-466-717-2
; Sequence 2, Application US/08466717
; Patent No. 5874234
```

```
Db 181 EDAQVVEDEEEDVEEEDVSGGDEDEEGYNDGVDGDEDEEELGEEERQKRK 238

RESULT 3
US-08-468-066-2
; Sequence 2, Application US/08468066
; Patent No. 5756676
; GENERAL INFORMATION:
; APPLICANT: Pasternack, Gary R.
; APPLICANT: Kuhaajda, Francis P.
; TITLE OF INVENTION: No. 5756676el Mammalian Protein Associated With
; TITLE OF INVENTION: Uncontrolled Cell Division
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington, D.C.
; STATE: District of Columbia
; COUNTRY: U.S.A.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,066
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,503
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Posorske Esq., Laurence H.
; REGISTRATION NUMBER: 34,698
; REFERENCE/DOCKET NUMBER: 1107.47218
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 508-9153
; TELEFAX: 202 508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-066-2

Query Match 82.3%; Score 1001; DB 1; Length 249;
Best Local Similarity 86.1%; Pred. No. 3.5e-88;
Matches 205; Conservative 7; Mismatches 22; Indels 4; Gaps 2;

Qy 1 MEMGRIHSELRNAPSVDKELALDNSRNEKLEALTDFEFLSKINGLTSIDL 60
Db 1 MEMGRIHLELRNRTSDVKELVLDNSRNEKLEGLTDFEFLSTINVGLTSIDL 60
Qy 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENKSIDL 116
Db 61 PKLKLKLELSDNRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKOLENKSIDL 120
Qy 117 FNCVTLNDYGNVFKLLQLTLYLSDCYWDHKEAPYSDIEHVEGLDDEEGEHEEYD 176
Db 121 FNCVTLNDYGNVFKLLQLTLYLSDCYWDHKEAPYSDIEHVEGLDDEEGEHEEYD 180
Qy 177 EDAQVVEDEEEDVEEEDVSGGDEDEEGYNDGVDGDEDEEELGEEERQKRK 234
Db 181 EDAQVVEDEEEDVEEEDVSGGDEDEEGYNDGVDGDEDEEELGEEERQKRK 238

RESULT 4
US-08-466-717-2
; Sequence 2, Application US/08466717
; Patent No. 5874234
```

**GENERAL INFORMATION:**

APPLICANT: Pasternack, Gary R.  
APPLICANT: Kuhajda, Francis P.  
TITLE OF INVENTION: NO. 5874234el Mammalian Protein Associated With  
TITLE OF INVENTION: Uncontrolled Cell Division  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, N.W.  
CITY: Washington, D.C.  
STATE: District of Columbia  
COUNTRY: U.S.A.  
ZIP: 20001

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,717
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,503
;

```

; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Posorske Esq., Laurence H.  
 ; REGISTRATION NUMBER: 34,698  
 ; REFERENCE/DOCKET NUMBER: 1107.47218  
 ; TELECOMMUNICATION INFORMATION:

```

; ; TELEPHONE: 202 508-9153
; ; TELEFAX: 202 508-9299
; ; INFORMATION FOR SEQ ID NO: 2:
; ;
; ; SEQUENCE CHARACTERISTICS:
; ;     LENGTH: 249 amino acids
; ;     TYPE: amino acid
; ;     TOPOLOGY: linear
; ;     MOLECULE TYPE: protein
US-08-466-717-2

```

|                           |       |                    |        |             |
|---------------------------|-------|--------------------|--------|-------------|
| Query Match               | 82.3% | Score 1001;        | DB 2;  | Length 249; |
| Best Local Similarity     | 86.1% | Pred. No. 3.5e-88; |        |             |
| Matches 205; Conservative |       | 7: Mismatches 22;  | Indels |             |

|    |     |  |     |
|----|-----|--|-----|
| QY | 1   | MEMGRRIHSELURNRAPSDVKELADNRSNEGKLEALTDFFELEFLSKINGGLTSIDL  | 60  |
|    |     |  |     |
| Db | 1   | MEMGRRIHSELURNRTPSDVKELVDNRSNEGKLEGLTDFEFLSTINVLGTSIANL    | 60  |
|    |     |  |     |
| QY | 61  | PKL-KLRKLEL--RVSGGLEVLAEKCPNTHLYLSGNKIKDLSITEPLKULENKLSIDL | 116 |
|    |     |  |     |
| Db | 61  | PKLNKLRKLELSDNRVSGGLEVLAEKCPNTHLSGNKIKDLSITEPLKULENKLSIDL  | 120 |
|    |     |  |     |
| QY | 117 | FNCEVTNLNDYGENVFKLLQLLYLDSQYWDHKEAPYSOTEDHVGLEDDEEGERHEEYD | 176 |
|    |     |  |     |
| Db | 121 | FNCEVTNLNDYRNVFKLLPQLLYLDGYDRDKAEPDPAEGYVGLDDEEDEDEREYD    | 180 |
|    |     |  |     |
| QY | 177 | EDAQVWDEEGEREEEGEEDYSGGDDEDEBYNDGVEDGDEEELGEERGOGRK        | 234 |
|    |     |  |     |
| Db | 181 | EDAQVWDEEDEEGEREEDYSGGEEDEBYNDGVEDDDEEELGEERGOGRK          | 238 |
|    |     |  |     |

## RESULT 5

US-08-766-738-4  
; Sequence 4, Application US/08766738  
; Patent No. 5916749  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Fastseq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/766,738  
FILING DATE: Herewith  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCES/DOCKET NUMBER: PP-0177 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:

```

/ ILLUM:
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 249 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GenBank
/ CLONE: 403007
/ US-08-766-738-4

```

Query Match 82.3%; Score 1001; DB 2; Length 249;  
Best Local Similarity 86.1%; Pred. No. 3.5e-88;  
Matches 205; Conservative 7; Mismatches 22; Indels

|    |     |   |     |
|----|-----|---|-----|
| QY | 1   | MEMGRIIHSJETURNRAPSVDKELALDNRSNEGKLEALTFEFEELFLSKINGLTSIDL  | 60  |
| Db | 1   | MEMGRIIHSJETURNRAPSVDKELALDNRSNEGKLEALTFEFEELFLSKINGLTSIDL  | 60  |
| QY | 61  | PKL-KLRKLEL--RVSGGLEVLAEKCPNTHLYLSGNKTKDLSTIEPKQLLENKLSLDL  | 116 |
| Db | 61  | PKLKNLKLLESDNRVSGGLEVLAEKCPNTHNLGNKTKDLSTIEPKQLLENKLSLDL    | 120 |
| QY | 117 | FNCVETNLNDYGENVFKLLQLTLDSCYWDHKEAPYSOTIEDHVEGLDDEEAGEHEEYD  | 176 |
| Db | 121 | FNCVETNLNDYRENVFKLLPQLTLYDGYDRDDKEAPDSOAEGYVEGLDDEEAGEHEEYD | 180 |
| QY | 177 | EDAQVVEDEEGEEEEEAGEEDYSGDDEEBGYNDGEVDGDDEDEELGEEERGGKRR     | 234 |
| Db | 181 | EDAQVVEDEDEDEEGEEEDYSGDEEEDGYNDGEVDDEDEDEELGEEERGGKRR       | 238 |

## RESULT 6

US-08-466-743-2  
; Sequence 2, Application US/08466743  
; Patent No. 6040173  
; GENERAL INFORMATION:  
; APPLICANT: Pasternack, Gary R.  
; APPLICANT: Kuhajda, Francis P.  
; TITLE OF INVENTION: No. 6040173el Mammalian Protein Associated With  
; TITLE OF INVENTION: Uncontrolled Cell Division  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner, Birch, McKie & Beckett  
; STREET: 1001 G Street, N.W.  
; CITY: Washington, D.C.  
; STATE: District of Columbia  
; COUNTRY: U.S.A.  
; ZIP: 20001



REGISTRATION NUMBER: 19,090  
REFERENCE/DOCKET NUMBER: 1107.51507  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 508-9153  
TELEFAX: 202 508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 249 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-12414-2

Query Match 82.3%; Score 1001; DB 5; Length 249;  
Best Local Similarity 86.1%; Pred. No. 3.5e-98;  
Matches 205; Conservative 7; Mismatches 22; Indels 4; Gaps 2;

QY 1 MEMGRRTHSELNRAPSDVKELALDNRSGKLEALTDFFEEFLSKINGLTSIDL 60  
DB 1 MEMGRRTHSELNRTPSDVKELVLDNRSGKLEGLTDEFEELFSLINVGLTSIANL 60  
QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTHLYLSGNKIKDLSLTPLEPKOLENLSL 116  
DB 61 PKLKLKLELSNVRSGGLEVLAEKCPNLTHLYLSGNKIKDLSLTPLEPKOLENLSL 120  
QY 117 FNCVNTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYD 176  
DB 121 FNCVNTNLNDYRENVFKLLPOLTYLDGYDDKAPSDAEAGYVEGLDDEEEDDEEYD 180  
QY 177 EDAQVVEDEGEDEEEDVSGGDEDEEGYNDGEVDGEDEEELGEERGQKRK 234  
DB 181 EDAQVVEDEDEDEEEDVSGEEDDEEGYNDGEVDGEDEEELGEERGQKRK 238

## RESULT 9

US-08-766-738-3  
; Sequence 3, Application US/08766738  
; Patent No. 5916749  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/766,738  
; FILING DATE: Herewith  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0177 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 251 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1498225  
US-08-766-738-3

Query Match 56.2%; Score 683.5; DB 2; Length 251;  
Best Local Similarity 61.0%; Pred. No. 8.7e-58;  
Matches 152; Conservative 34; Mismatches 40; Indels 23; Gaps 8;

QY 1 MEMGRRTHSELNRAPSDVKELALDNRSGKLEALTDFFEEFLSKINGLTSIDL 60  
DB 1 MDKRRTHLELRNRTPAAVRELVLDNRSGKLEGLTAEFVNLEFSLINVGLSVSNL 60  
QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTHLYLSGNKIKDLSLTPLEPKOLENLSL 116  
DB 61 PKLPLKLELSNRFGLDMLAEKLPNLTHLYLSGNKIKDLSLTPLEPKOLENLSL 120  
QY 117 FNCVNTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEY 175  
DB 121 FNCVNTNLNDYRENVFKLLPOLTYLDGYDDKAPSDAE--VDGVDDEEEDDEE 178  
QY 176 EDAQVVEDEGEDEE---EEGEEEDVSG-----GDEDEEGYNDGEVDGEDEEELG 225  
DB 179 DEDE-----DEGEEEDDEDEDDDEGDEDEDEDEDEDEDEDEDEDEDEE-E 232  
QY 226 EEERGGKRK 234  
DB 233 EGGKEKRK 241

## RESULT 10

US-09-262-610-3  
; Sequence 3, Application US/09262610  
; Patent No. 6428949  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/262,610  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/766,738  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0177 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 251 amino acids  
; TYPE: amino acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1498225
; US-09-262-610-3

Query Match
Best Local Similarity 56.2%; Score 683.5; DB 4; Length 251;
Matches 152; Conservative 34; Mismatches 40; Indels 23; Gaps 8;

QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNECKLEALTDEFELEFLSKINGLTSIDL 60
DB 1 MDKRRHLELRNRTPAAVRELVDNCKSNDKIGLTAEFVNFELSLINVLISVNL 60
QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTLYLTHLNSGNKIKDLSITIEPLKQLENKLSLDL 116
DB 61 PKLPKLLKLELSENRIFFGLDMLAEKLPNLTNLSGNKLDISTIEPLKLECLKSLDL 120
QY 117 FNCVNTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSIEDHVGLEDDEEHEEY 175
DB 121 FNCVNTNLNDYRESVFKLLPQLTYLDGYDRDQAPDSAE--VDGVDEEEDDEGEDEE 178
QY 176 DEDAQVVEDEGEDEE---EEGEEDYSG-----GDEDEEGYNDGEVDGEDEELG 225
DB 179 DED-----DEGGEEDFDEDEDEDEVDGEDEDEVESEEEFGLDEDEDEDEE-E 232
QY 226 EERGQKRK 234
DB 233 EGGKGEKRK 241

RESULT 11
US-08-766-738-1
; Sequence 1, Application US/08766738
; Patent No. 5916749
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,738
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0177 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 251 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: 1813361
; US-08-766-738-1

Query Match
Best Local Similarity 55.6%; Score 676.5; DB 2; Length 251;
Matches 151; Conservative 34; Mismatches 41; Indels 23; Gaps 8;

QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNECKLEALTDEFELEFLSKINGLTSIDL 60
DB 1 MDKRRHLELRNRTPAAVRELVDNCKSNDKIGLTAEFVNFELSLINVLISVNL 60
QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTLYLTHLNSGNKIKDLSITIEPLKQLENKLSLDL 116
DB 61 PKLPKLLKLELSENRIFFGLDMLAEKLPNLTNLSGNKLDISTIEPLKLECLKSLDL 120
QY 117 FNCVNTNLNDYGENVFKLLQLTYLDSCYWDHKEAPYSIEDHVGLEDDEEHEEY 175
DB 121 FNCVNTNLNDYRESVFKLLPQLTYLDGYDRDQAPDSAE--VDGVDEEEDDEGEDEE 178
QY 176 DEDAQVVEDEGEDEE---EEGEEDYSG-----GDEDEEGYNDGEVDGEDEELG 225
DB 179 DED-----DEGGEEDFDEDEDEDEVDGEDEDEVESEEEFGLDEDEDEE-E 232
QY 226 EERGQKRK 234
DB 233 EGGKGEKRK 241

RESULT 12
US-09-262-610-1
; Sequence 1, Application US/09262610
; Patent No. 6428949
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,610
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/766,738
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0177 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 251 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```



IMMEDIATE SOURCE:  
LIBRARY: Consensus  
CLONE: 1813361  
US-09-262-610-1

Query Match 55.6%; Score 676.5; DB 4; Length 251;  
Best Local Similarity 60.6%; Pred. No. 4.1e-57;  
Matches 151; Conservative 34; Mismatches 41; Indels 23; Gaps 8;  
QY 1 MEMGRIHSELRNAPSVDVKELADNRSNGKLEALTDFEELEFLSKINGGITSIDL 60  
DQ 1 MDKRRIRHLELRNTPAAVRELVDNCKSNCKIEGLTAETVNLNGLISVNL 60  
QY 61 PKL-KRLKLEL---RVSGGLEVLAEKCPNLTHTLYLSGNKIKDLSTIEPLKQLENLKSLDL 116  
DQ 61 PKLPLKLLKLESENRIFGGLDMLAEKLPNLTHTLYLSGNKIKDLSTIEPLKQLENLKSLDL 120  
QY 117 FNCVNTNLDYGENVFKLLQLTYLDSYWDHKEAPYSIDIEDHVEGLD-DEEGERHEEY 175  
DQ 121 FNCVNTNLDYRESVFKLLPQLTYLDGYDREDQAPSDAE--VDGYDXXEEDGEGED 178  
QY 176 DEDAQVDEGEER---EEGEEDVSG-----GDEDEGYNDGEVDGEDEELG 225  
DQ 179 DED-----DEGEDEEEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEE-E 232  
QY 226 EEEGQKRK 234  
DQ 233 EGGGKRK 241

RESULT 13  
US-08-466-603-5  
Sequence 5, Application US/08466603  
Patent No. 5726018  
GENERAL INFORMATION:  
APPLICANT: Pasternack, Gary R.  
APPLICANT: Kujajda, Francis P.  
TITLE OF INVENTION: No. 5726018el Mammalian Protein Associated With  
TITLE OF INVENTION: Uncontrolled Cell Division  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, N.W.  
CITY: Washington, D.C.  
STATE: District of Columbia  
COUNTRY: U.S.A.  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,603  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,503  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Posorske Esq., Laurence H.  
REGISTRATION NUMBER: 34,698  
REFERENCE/DOCKET NUMBER: 1107.47218  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 508-9153  
TELEFAX: 202 508-9299  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 182 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-603-5

Query Match 53.3%; Score 648.5; DB 1; Length 182;  
Best Local Similarity 77.5%; Pred. No. 1.2e-54;  
Matches 134; Conservative 11; Mismatches 23; Indels 5; Gaps 2;  
QY 65 LRKLEL---RVSGGLEVLAEKCPNLTHTLYLSGNKIKDLSTIEPLKQLENLKSLDLFNCV 121  
DQ 1 VKKLESENRIKSDLEVLAEKCPNLTHTLYLSGNKIKDLSTIEPLKQLENLKSLDLFNCV 60  
QY 122 TNLNDYGENVFKLLQLTYLDSYWDHKEAPYSIDIEDHVEGLDDEEGERHEEYDQAV 181  
DQ 61 TNLNAYRENVFKLLPQWYLDGYDRNKEAPSDVGEYVE--DDDEDEDEEYDEYDQAV 118  
QY 182 VEDEGEEREEGEEDVSGDEDEDEGYNDGEVDGEDEEELGEEERQKRK 234  
DQ 119 VEDEEEVEEVEEEDVSGDEDEDEGYNDGEVDGEDEEELGEEERQKRK 171

RESULT 14  
US-08-314-503A-5  
Sequence 5, Application US/08314503A  
Patent No. 5734022  
GENERAL INFORMATION:  
APPLICANT: Pasternack, Gary R.  
APPLICANT: Kujajda, Francis P.  
TITLE OF INVENTION: No. 5734022el Mammalian Protein Associated With  
TITLE OF INVENTION: Uncontrolled Cell Division  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner, Birch, McKie & Beckett  
STREET: 1001 G Street, N.W.  
CITY: Washington, D.C.  
STATE: District of Columbia  
COUNTRY: U.S.A.  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/314,503A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Posorske Esq., Laurence H.  
REGISTRATION NUMBER: 34,698  
REFERENCE/DOCKET NUMBER: 1107.47218  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 508-9153  
TELEFAX: 202 508-9299  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 182 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-314-503A-5

Query Match 53.3%; Score 648.5; DB 1; Length 182;  
Best Local Similarity 77.5%; Pred. No. 1.2e-54;  
Matches 134; Conservative 11; Mismatches 23; Indels 5; Gaps 2;  
QY 65 LRKLEL---RVSGGLEVLAEKCPNLTHTLYLSGNKIKDLSTIEPLKQLENLKSLDLFNCV 121  
DQ 1 VKKLESENRIKSDLEVLAEKCPNLTHTLYLSGNKIKDLSTIEPLKQLENLKSLDLFNCV 60  
QY 122 TNLNDYGENVFKLLQLTYLDSYWDHKEAPYSIDIEDHVEGLDDEEGERHEEYDQAV 181  
DQ 61 TNLNAYRENVFKLLPQWYLDGYDRNKEAPSDVGEYVE--DDDEDEDEEYDEYDQAV 118  
QY 182 VEDEGEEREEGEEDVSGDEDEDEGYNDGEVDGEDEEELGEEERQKRK 234

Db 119 VEDEEEVEEERGEDVSGEEDEDEEGYNDGEVDDEDEEAGEEGSQRRK 171

RESULT 15

US-08-468-066-5  
; Sequence 5, Application US/08468066  
; Patent No. 5756676  
; GENERAL INFORMATION:  
; APPLICANT: Pasternack, Gary R.  
; APPLICANT: Kuhajda, Francis P.  
; TITLE OF INVENTION: No. 5756676el Mammalian Protein Associated With  
; TITLE OF INVENTION: Uncontrolled Cell Division  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner, Birch, McKie & Beckett  
; STREET: 1001 G Street, N.W.  
; CITY: Washington, D.C.  
; STATE: District of Columbia  
; COUNTRY: U.S.A.  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,066  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,503  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Posorske Esq., Laurence H.  
; REGISTRATION NUMBER: 34,698  
; REFERENCE/DOCKET NUMBER: 1107.47218  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202 508-9153  
; TELEFAX: 202 508-9299  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 182 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-468-066-5

Query Match 53.3%; Score 648.5; DB 1; Length 182;  
Best Local Similarity 77.5%; Pred. No. 1.2e-54;  
Matches 134; Conservative 11; Mismatches 23; Indels 5; Gaps 2;  
QY 65 LRKLEL--RVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKQLENLKSLDLFNCEV 121  
Db 1 VKKLELSERNISGDLEVLAEKCPNLKHLNLSGNKIKDLSTIEPLKKLENLKSLDLFNCEV 60  
QY 122 TNLNDYGENVFILLQLTYLDSCYWDHKEAPYSDIEDHVGLEDDEEGEHEEYDEDAQV 181  
Db 61 TNLNAYRENVFILLPQVMYLDGYDRDNKEAPDSVGEYVE--DDDEDEDEEYDEYQAL 118  
QY 182 VEDEEEVEEERGEDVSGDEEEDGEGYNDGEVDGEDEELGEEERQRRK 234  
Db 119 VEDEEEVEEERGEDVSGEEDEDEEGYNDGEVDDEDEEAGEEGSQRRK 171

Search completed: July 7, 2003, 14:59:31  
Job time : 36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2003, 14:56:03 ; Search time 30 Seconds  
(without alignments)  
896.945 Million cell updates/sec

Title: US-09-591-500A-4

Perfect score: 1216

Sequence: 1 MEMGRIHSELNRAPSDVK.....VDGEDEELGEEERGQK 234

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA.\*

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 1001  | 82.3        | 249    | 9  | US-10-213-700-4   |
| 2          | 683.5 | 56.2        | 251    | 9  | US-10-213-700-3   |
| 3          | 676.5 | 55.6        | 251    | 9  | US-10-213-700-1   |
| 4          | 175   | 14.4        | 76     | 9  | US-10-101-487-36  |
| 5          | 173   | 14.2        | 180    | 9  | US-10-101-487-116 |
| 6          | 173   | 14.2        | 197    | 9  | US-10-101-487-51  |
| 7          | 171   | 14.1        | 197    | 9  | US-10-101-487-114 |
| 8          | 170   | 14.0        | 200    | 9  | US-10-101-487-53  |
| 9          | 169   | 13.9        | 179    | 9  | US-10-101-487-107 |
| 10         | 169   | 13.9        | 181    | 9  | US-10-101-487-45  |
| 11         | 166   | 13.7        | 176    | 9  | US-10-101-487-70  |
| 12         | 166   | 13.7        | 177    | 9  | US-10-101-487-48  |
| 13         | 166   | 13.7        | 177    | 9  | US-10-101-487-115 |
| 14         | 166   | 13.7        | 179    | 9  | US-10-101-487-46  |
| 15         | 166   | 13.7        | 186    | 9  | US-10-101-487-44  |
| 16         | 166   | 13.7        | 187    | 9  | US-10-101-487-50  |
| 17         | 166   | 13.7        | 191    | 9  | US-10-101-487-81  |
| 18         | 166   | 13.7        | 198    | 9  | US-10-101-487-42  |
| 19         | 165   | 13.6        | 174    | 9  | US-10-101-487-72  |

165 13.6 175 9 US-10-101-487-57  
165 13.6 176 9 US-10-101-487-56  
165 13.6 240 9 US-10-101-487-75  
165 13.6 350 9 US-10-101-487-58  
160 13.2 587 9 US-09-893-519A-42  
156 12.8 376 9 US-10-156-761-9889  
150.5 12.4 740 9 US-10-128-714-3032  
150.5 12.4 740 9 US-10-128-714-3032  
147.5 12.1 119 9 US-09-864-761-44441  
147 12.1 114 9 US-09-864-761-36168  
146.5 12.0 440 9 US-10-025-380-1059  
146.5 12.0 440 10 US-09-922-217-1059  
146.5 12.0 440 10 US-09-833-263-1059  
144.5 11.9 272 10 US-09-922-261-186  
142.5 11.7 62 10 US-09-864-761-35828  
141.5 11.6 58 9 US-10-101-487-43  
141 11.6 89 10 US-09-864-761-44120  
140.5 11.6 714 10 US-09-978-242-3  
136.5 11.2 231 10 US-09-922-261-194  
136.5 11.2 232 10 US-09-922-261-192  
136.5 11.2 238 10 US-09-922-261-190  
136.5 11.2 257 10 US-09-922-261-188  
135.5 11.1 743 9 US-10-087-464-53  
133 10.9 534 9 US-09-893-519A-37  
133 10.9 1504 9 US-09-932-145-7  
131.5 10.8 376 9 US-10-156-761-9888

#### ALIGNMENTS

#### RESULT 1

US-10-213-700-4

Sequence 4, Application US/10213700

Publication No. US20030022332A1

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/213,700

FILING DATE: 06-Aug-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/766,738

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PP-0177 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 249 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

Sequence 57, Appl  
Sequence 56, Appl  
Sequence 75, Appl  
Sequence 58, Appl  
Sequence 42, Appl  
Sequence 9889, Ap  
Sequence 3032, Ap  
Sequence 8032, Ap  
Sequence 4444, A  
Sequence 36168, A  
Sequence 1059, Ap  
Sequence 1059, Ap  
Sequence 1059, Ap  
Sequence 186, App  
Sequence 35828, A  
Sequence 43, Appl  
Sequence 44120, A  
Sequence 3, Appl  
Sequence 194, App  
Sequence 192, App  
Sequence 190, App  
Sequence 188, App  
Sequence 53, Appl  
Sequence 37, Appl  
Sequence 7, Appl  
Sequence 9888, Ap

```

; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 403007
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-213-700-4

Query Match      82.3%; Score 1001; DB 9; Length 249;
Best Local Similarity 86.1%; Pred. No. 2 4e-65;
Matches 205; Conservative 7; Mismatches 22; Indels 4; Gaps 2;

Qy      1 MEMGRRHSELNRNAPSVDKVELALDNRSGNEKGLPALTDFEPELEFLSKINGLTSIDL 60
Db      1 MEMGRRHLELRNRTSDYKVELVDNRSGNEKLEGLTDFEPELEFLSTINVLGTSTANL 60

Qy      61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTHTLYLSGNKIKDLSLTIPLKOLENKLSDL 116
Db      61 PKLNKLLKLELSDNRVSGGLEVLAEKCPNLTHTLNSGNKIKDLSLTIPLKLENLKLSL 120

Qy      117 FNCVETNLNDYGENVFKLLQLLYLQSCYWDHKAPYSDIEDRVHVGLEDDEGEHEEYD 176
Db      121 FNCVETNLNDYRENVFKLLPQLYLYLQYDRDKKAPDSDAEYVVGLEDDEEYD 180

Qy      177 EDAQVWDEGEDEEEDYSGDDEDEEYNDYGVDEDEEELGEEERGOKR 234
Db      191 FCAQVWDETFDFEPEFEDYSGFEEDYNGYNDYGVDEDEEELGEEERGOKR 238

```

```

1  RESULT 2
2  US-10-213-700-3
3  ; Sequence 3, Application US/10213700
4  ; Publication No. US20030022332A1
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Bandman, Olga
7  ; Goli, Surya K.
8  ; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
9  ; NUMBER OF SEQUENCES: 4
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
12 ; STREET: 3174 Porter Drive
13 ; CITY: Palo Alto
14 ; STATE: CA
15 ; COUNTRY: USA
16 ; ZIP: 94304
17 ; COMPUTER READABLE FORM:
18 ; MEDIUM TYPE: Diskette
19 ; COMPUTER: IBM Compatible
20 ; OPERATING SYSTEM: DOS
21 ; SOFTWARE: FastSeq for Windows Version 2.0
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: US/10/213,700
24 ; FILING DATE: 06-Aug-2002
25 ; CLASSIFICATION: <Unknown>
26 ; PRIOR APPLICATION DATA:
27 ; APPLICATION NUMBER: US/08/766,738
28 ; FILING DATE: <Unknown>
29 ; ATTORNEY/AGENT INFORMATION:
30 ; NAME: Billings, Lucy J.
31 ; REGISTRATION NUMBER: 36,749
32 ; REFERENCE/DOCKET NUMBER: PF-0177 US
33 ; TELECOMMUNICATION INFORMATION:
34 ; TELEPHONE: 415-855-0555
35 ; TELEFAX: 415-845-4166
36 ; TELEX: <Unknown>
37 ; INFORMATION FOR SEQ ID NO: 3:
38 ; SEQUENCE CHARACTERISTICS:
39 ; LENGTH: 251 amino acids
40 ; TYPE: amino acid
41 ; STRANDEDNESS: single
42 ; TOPOLOGY: linear
43 ; IMMEDIATE SOURCE:
44 ; LIBRARY: GenBank
45 ; CLONE: 1498225
46 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```



```
; OTHER INFORMATION: protein
US-10-101-487-53

Query Match      14.0%; Score 170; DB 9; Length 200;
Best Local Similarity 44.2%; Pred. No. 2.9e-05;
Matches 38; Conservative 16; Mismatches 32; Indels 0; Gaps 0;

QY 147 DHKEAPYSIDIEHVGLEDDEEGEHEEYDEDAQVVEDEGEDEEEDYSGGDEED 206
Db 105 EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 164
QY 207 EEGYNDGVDGEDEDEELGEEERGOK 232
Db 165 EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 190

RESULT 9
US-10-101-487-107
; Sequence 107, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOQUIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 107
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-107

Query Match      13.9%; Score 169; DB 9; Length 179;
Best Local Similarity 41.1%; Pred. No. 3e-05;
Matches 37; Conservative 19; Mismatches 34; Indels 0; Gaps 0;

QY 145 YWDHKEAPYSIDIEHVGLEDDEEGEHEEYDEDAQVVEDEGEDEEEDYSGGDE 204
Db 1 YFQGEDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEE 60
QY 205 EDEGYNDGVDGEDEDEELGEEERGOKR 234
Db 61 EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 90

RESULT 10
US-10-101-487-45
; Sequence 45, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOQUIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
```

```
QY 207 EEGYNDGVDGEDEDEELGEEERGOKR 234
Db 130 EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 157

RESULT 7
US-10-101-487-114
; Sequence 114, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOQUIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 114
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-114

Query Match      14.1%; Score 171; DB 9; Length 197;
Best Local Similarity 39.8%; Pred. No. 2.4e-05;
Matches 37; Conservative 21; Mismatches 35; Indels 0; Gaps 0;

QY 142 DSCYWDHKEAPYSIDIEHVGLEDDEEGEHEEYDEDAQVVEDEGEDEEEDYSG 201
Db 12 ENLYFQGEDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEEDDEEED 71
QY 202 GDEDEGYNDGVDGEDEDEELGEEERGOKR 234
Db 72 EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 104

RESULT 8
US-10-101-487-53
; Sequence 53, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOQUIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic fusion
```

US-10-101-487-115

|                       |                  |                  |           |             |
|-----------------------|------------------|------------------|-----------|-------------|
| Query Match           | 13.7%            | Score 166;       | DB 9;     | Length 177; |
| Best Local Similarity | 43.4%;           | Pred. No. 5e-05; |           |             |
| Matches 36;           | Conservative 17; | Mismatches 30;   | Indels 0; | Gaps        |

QY 152 PYSIEDHVEGLDDEEGEHEEYDEDAQVVEDEEGEEEEEEDVSGGDEDEGYN 21





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2003, 14:54:17 ; Search time 40 seconds  
(without alignments)  
562.386 Million cell updates/sec

Title: US-09-591-500A-4  
Perfect score: 1216  
Sequence: 1 MEMGRRHSELNRAPSDVK.....VDGEDEELGEERQKRR 234  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID       | Description        |
|------------|-------|---------------|--------|----------|--------------------|
| 1          | 1001  | 82.3          | 249    | 2 S43309 | probable HLA class |
| 2          | 883   | 72.6          | 247    | 2 I59334 | leucine-rich acid  |
| 3          | 667   | 54.9          | 272    | 2 JC7357 | 31K proliferation  |
| 4          | 526.5 | 43.3          | 134    | 2 S36375 | PHAP1 protein - hu |
| 5          | 294.5 | 24.2          | 447    | 2 T46146 | hypothetical prote |
| 6          | 243   | 20.0          | 229    | 2 T21714 | hypothetical prote |
| 7          | 230   | 18.9          | 225    | 2 T34456 | hypothetical prote |
| 8          | 177.5 | 14.6          | 556    | 2 B46024 | neurofilament-L su |
| 9          | 176   | 14.5          | 409    | 2 E86336 | hypothetical prote |
| 10         | 174.5 | 14.4          | 544    | 2 B44841 | low molecularweig  |
| 11         | 170.5 | 14.0          | 792    | 2 T42963 | hypothetical prote |
| 12         | 167.5 | 13.8          | 590    | 2 A04037 | glutamic acid-rich |
| 13         | 166.5 | 13.7          | 543    | 1 QFMSL  | hypothetical prote |
| 14         | 165   | 13.6          | 721    | 2 S29795 | hypothetical prote |
| 15         | 165   | 13.6          | 797    | 2 A36811 | probable cell divi |
| 16         | 165   | 13.6          | 896    | 2 T24169 | neurofilament trip |
| 17         | 165   | 13.6          | 906    | 2 T24166 | hypothetical prote |
| 18         | 164   | 13.5          | 1019   | 2 T40813 | hypothetical prote |
| 19         | 163.5 | 13.4          | 544    | 2 S07144 | probable cell divi |
| 20         | 163   | 13.4          | 678    | 2 A34514 | neurofilament trip |
| 21         | 162.5 | 13.4          | 364    | 1 EDBESP | glutamic acid-rich |
| 22         | 161   | 13.2          | 972    | 2 T49773 | related to actin-i |
| 23         | 160   | 13.2          | 587    | 2 JC5300 | Ran GTPase activat |
| 24         | 160   | 13.2          | 798    | 2 I50479 | neurofilament medi |
| 25         | 159.5 | 13.1          | 1300   | 2 T03166 | probable immediate |
| 26         | 158.5 | 13.0          | 636    | 2 T51893 | related to Che-1 p |
| 27         | 158   | 13.0          | 279    | 2 S30766 | ASF1 protein - yea |
| 28         | 157   | 12.9          | 630    | 2 S29796 | hypothetical prote |
| 29         | 156   | 12.8          | 81     | 2 T48398 | hypothetical prote |

|    |       |      |      |          |                    |
|----|-------|------|------|----------|--------------------|
| 30 | 156   | 12.8 | 1262 | 2 T33074 | hypothetical prote |
| 31 | 155   | 12.7 | 599  | 2 S18735 | centromere protein |
| 32 | 155   | 12.7 | 858  | 2 S15762 | neurofilament trip |
| 33 | 155   | 12.7 | 913  | 2 S52485 | neurofilament prot |
| 34 | 154   | 12.7 | 589  | 2 A36983 | RNAI homolog fugu  |
| 35 | 154   | 12.7 | 589  | 2 T52070 | RNAI protein homol |
| 36 | 153   | 12.6 | 548  | 1 QFFGL  | neurofilament trip |
| 37 | 151   | 12.4 | 916  | 2 A27864 | neurofilament trip |
| 38 | 150   | 12.3 | 606  | 2 S70358 | centromere protein |
| 39 | 150   | 12.3 | 670  | 2 T28391 | ORF MS230 hypothe  |
| 40 | 149.5 | 12.3 | 407  | 1 EDBEQ3 | immediate-early pr |
| 41 | 148.5 | 12.2 | 262  | 2 T41536 | yeast anti-silenci |
| 42 | 148.5 | 12.2 | 1132 | 2 T43483 | translation initia |
| 43 | 147.5 | 12.1 | 168  | 2 F84499 | hypothetical prote |
| 44 | 147.5 | 12.1 | 845  | 2 A45669 | neurofilament trip |
| 45 | 147   | 12.1 | 325  | 2 T18283 | hypothetical prote |

ALIGNMENTS

RESULT 1  
S43309  
probable HLA class II-associated protein PHAP1 - human  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 17-Nov-2000  
C:Accession: S43309; S37222  
R:Vaesen, M.; Barnikol-Watanabe, S.; Goetz, H.; Awml, L.A.; Cole, T.; Zimmermann, B.,  
Biol. Chem. Hoppe-Seyler 375, 113-126, 1994  
A:Title: Purification and characterization of two putative HLA class II associated pr  
A:Reference number: S43309; MUID:94250340; PMID:8192856  
A:Accession: S43309  
A:Molecule type: mRNA  
A:Residues: 1-249 <VAE>  
A:Cross-references: EMBL:X75090; NID:q403006; PIDN:CAA52981.1; PID:q403007  
C:Superfamily: Caenorhabditis elegans hypothetical protein T19H12.2  
C:Keywords: signal transduction

Query Match 82.3%; Score 1001; DB 2; Length 249;  
Best Local Similarity 86.1%; Pred. No. 2.7e-54;  
Matches 205; Conservative 7; Mismatches 22; Indels 4; Gaps 2;  
QY 1 MEMGRRHSELNRAPSDVKELALDNSRSGKLEALTDEFEELFLSKINGLTSDLS 60  
DB 1 MEMGRRHLELRNTPSDVKELVDNSRSGKLEGLTDEFEELFLSTINVLSTANL 60  
QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKSLDL 116  
DB 61 PKLNKAKLELSDNRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKSLDL 120  
QY 117 FNCVETNLDYGENVFKLLQLTYLDSYWDHKEAPYSDIEDHVEGLDDEEGEHEEYD 176  
DB 121 FNCVETNLDYRENVFKLLPOLTYLDSYWDHKEAPYSDIEDHVEGLDDEEGEHEEYD 180  
QY 177 EDAQVVEDEGEDEEEDVSGGDEDEEGVNDGEVDEGEELGEERQKRR 234  
DB 181 EDAQVVEDEGEDEEEDVSGGDEDEEGVNDGEVDEGEELGEERQKRR 238

RESULT 2  
I59334  
leucine-rich acidic nuclear protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 17-Nov-2000  
C:Accession: I59334  
R:Matsuoka, K.; Taoka, M.; Satozawa, N.; Nakayama, H.; Ichimura, T.; Takahashi, N.; Y  
Proc. Natl. Acad. Sci. U.S.A. 91, 9670-9674, 1994  
A:Title: A nuclear factor containing the leucine-rich repeats expressed in murine cer  
A:Reference number: I59334; MUID:95024022; PMID:7937870  
A:Accession: I59334  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-247 <RES>

```
A:Cross-references: GB:D32209; NID:g511663; PIDN:BAA0908.1; PID:g511664
C:Superfamily: Caenorhabditis elegans hypothetical protein T19H12.2

Query Match      72.6%; Score 883; DB 2; Length 247;
Best Local Similarity 77.3%; Pred. No. 4.1e-47;
Matches 184; Conservative 16; Mismatches 32; Indels 6; Gaps 3;

QY 1 MEMGRIHSELNRRAPSDVKELALDNRSGKLEALTDFEEFLSKINGLTSIDL 60
Db 1 MEMDKRIYELNRNTPSDVKELVLDNCRSIEGKIEGLTDFEEFLSTINVGITSINL 60

QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKOLENLSIDL 116
Db 61 PLNKLUKLULESENRIISGLVEIAEKCPNLKHLNLSGNKIKDLSTIEPLKALENLSIDL 120

QY 117 FNCVETNLNDYGENVFKLLQLTYLDSYWDHKEAPYSIDIEDHVEGLDDDEEGHEEYD 176
Db 121 FNCVETNLNAYRENVPKLLPQVMYLDGYDRNKEAPDSVGEYVE--DDDEDEDEYD 178

QY 177 EDQVVEDEGEDEEEDEEEDVSGDEDEEGYNDGEVDGEDEEELGEERQKRR 234
Db 179 EYAQLVEDEEEDDEEEDVSGDEEEDVSGYNDGEVDDEDEDAEESGQKRR 236

RESULT 3
JC7357
31K proliferation related acidic leucine-rich protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 17-Nov-2000
C:Accession: JC7357
R:Mutai, H.; Toyoshima, Y.; Sun, W.; Hattori, N.; Tanaka, S.; Shiota, K.
Biochem. Biophys. Res. Commun. 274, 427-433, 2000
A:Title: PAL31, a novel nuclear protein, expressed in the developing brain.
A:Reference number: JC7357
A:Accession: JC7357
A:Molecule type: mRNA
A:Residues: 1-272 <MUT>
A:Cross-references: DDBJ:AB025581
A:Comment: This protein, a nuclear protein, is involved in brain development through the
C:Genetics:
A:Gene: pal31
C:Superfamily: Caenorhabditis elegans hypothetical protein T19H12.2
C:Keywords: brain; nucleus; tandem repeat

Query Match      54.9%; Score 667; DB 2; Length 272;
Best Local Similarity 55.3%; Pred. No. 6.6e-34;
Matches 147; Conservative 39; Mismatches 45; Indels 34; Gaps 7;

QY 1 MEMGRIHSELNRRAPSDVKELALDNRSGKLEALTDFEEFLSKINGLTSIDL 60
Db 1 MDKRRITHLELRNRTPAAVQELVLDNCKANDGKIEGLTDFEYNLEFLSLINVLFSVSDL 60

QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKOLENLSIDL 116
Db 61 PKLPLKALKLESENRIFFGGDLRAELPSLTHLYLSGNKIKDLSTIEPLKLUKLUKLSIDL 120

QY 117 FNCVETNLNDYGENVFKLLQLTYLDSYWDHKEAPYSIDIE-----DHVEGLDD 165
Db 121 FGCETNRSYRETVFLLPQLSYLDGYDREDEAPDSVDVSEAPDSGDGVDK 180

QY 166 EEEGE---HHEEYDEDAQVVEDEEEEEEDEEEDVSGDEED-----BEGYNDGEV 215
Db 181 EEEDEGEDEEEDDG-----EEDEDEDEDEDEDEDEDEDEDEVEGDEEFGHDGEV 237

QY 216 DGEDEEELGEE-----RGQKRR 234
Db 238 DEDEDEDEDEDEEESGKARR 262

RESULT 4
S36375
PHAPI protein - human (fragments)
C:Species: Homo sapiens (man)
```

```
RESULT 6
T21714
hypothetical protein F33H2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C:Accession: T21714
R:Cottage, A.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19463
A:Accession: T21714
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-229 <WIL>
A:Cross-references: EMBL:Z81526; PIDN:CA804265.1; GSPDB:GN00019; CESP:F33H2.3
A:Experimental source: clone F33H2
C:Genetics:
A:Gene: CESP:F33H2.3
A:Map position: 1
A:Introns: 199/3
C:Superfamily: Caenorhabditis elegans hypothetical protein T19H12.2

Query Match 20.0%; Score 243; DB 2; Length 229;
Best Local Similarity 31.9%; Pred. No. 3.7e-08;
Matches 75; Conservative 40; Mismatches 88; Indels 32; Gaps 9;

QY 10 ELRRAPSDYKELALDNRSGNECKLEALTDEFEFEFLSKINGLTSISDLPR-----L 63
DB 11 ELRRDPATVDTFLDN--AEDGQIGGLTDLINLEMLSMVKGLTTLGFPPTLPALTYL 68
QY 64 KLRLELRVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKOLENKLKSLDLF-NCEVT 122
DB 69 DISDNQLGDNASDFVLVKNAPDLKKITLANKL-SLDNLRCLKVLNPLFELDSLNPISG 127
QY 123 NLNDYGENVFKLLQLTYLDSYWDHKEAPYSIEDIHDV--EGLDDEEGEHEEE-----Y 175
DB 128 LLEDYRKFMFEMIPSLKILDCDVGEE-----VEEFAGEGGEDSEEGSGDDEGPGLSY 182
QY 176 DEDAQVDEDEGESEEEEDVSGGDEDEEGYNDGEVDGEE--DEEELGEE 228
DB 183 LEKSQSDDETDYAPE-----GGDAEPRTGKRGASDNGEEDPNKKAAGDDE 229

RESULT 7
T34456
hypothetical protein T19H12.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 08-Sep-2000
C:Accession: T34456
R:Davidson, S.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid T19H12.
A:Reference number: Z21528
A:Accession: T34456
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-225 <DAV>
A:Cross-references: EMBL:U97009; PIDN:AAC69027.1; GSPDB:GN00023; CESP:T19H12.2
A:Experimental source: strain Bristol N2; clone T19H12
C:Genetics:
A:Gene: CESP:T19H12.2
A:Map position: 5
A:Introns: 41/3; 199/3
C:Superfamily: Caenorhabditis elegans hypothetical protein T19H12.2

Query Match 18.9%; Score 230; DB 2; Length 225;
Best Local Similarity 30.0%; Pred. No. 2.3e-07;
Matches 71; Conservative 34; Mismatches 86; Indels 46; Gaps 7;

QY 9 SELRRAPSDYKELALDNRSGNECKLEALTDEFEFEFLSKINGLTSISDLP-----K 62
DB 7 SELRGREPVTDTFLDNTQG--GVIGGINEKLTKELLSMVKGLTTLGKMPVLPALNY 64
```

```
QY 63 LKLRKLELRVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKOLENKLKSLDLF-NCEV 121
DB 65 LDLSNELGDDASDFVLKCAPEIKKITLSGNRL-TLDNVRTKMLPNLMELDLNNSSL 123
QY 122 TNLNDYGENVFKLLQLTYLDSYWDHKEAPYSIEDIHDVGLDDEEGEHEEYDEDAQV 181
DB 124 GLLDYRVKMFEMIPSLKILDCG-----DVGDEE 152
QY 182 VEDE----EGEEEGE--EDVSGGDEDEEGYNDGEVDGEEDEEELGEEERGQKR 233
DB 153 VEEFAAGECAEDSDGSDDEGPGLSYLKNSQFSDDETDDYVVPAGDAETRGAKR 209

RESULT 8
B46024
neurofilament-L subunit - quail
C:Species: Coturnix coturnix (quail)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 12-Apr-1995
C:Accession: B46024
R:Ohara, O.; Gahara, Y.; Miyake, T.; Teraoka, H.; Kitamura, T.
J. Cell Biol. 121, 387-395, 1993
A:title: Neurofilament deficiency in quail caused by nonsense mutation in neurofilament
A:Reference number: A46024; MUID:93224534; PMID:8468353
A:Accession: B46024
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-556 <OHA>
A:Experimental source: subsp. japonica, TKP
A>Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:129455, NCBIP:129482)
C:Superfamily: cytoskeletal keratin

Query Match 14.6%; Score 177.5; DB 2; Length 556;
Best Local Similarity 26.5%; Pred. No. 0.00096;
Matches 65; Conservative 38; Mismatches 91; Indels 51; Gaps 9;

QY 20 KELALDNRSGKLEALTDEFEFEFLSKINGLTSISDLPLKLRKLELRVSGGLEVL 79
DB 317 KYLEIATR--GMNEALEKQLELE--EKQSANISALQD----YINKLENLRTTKSM 367
QY 80 AEKCPNLTHLYLSGNKIKDLSTIEPLKOLENKLKSLDLFNCVETNUN----- 125
DB 368 AR-----YL--KEYQLLVKALDIEIAAYRKLLEGEETRLSFTSVGVTSGYTQT 417
QY 126 --DYGENVFKLLQLTYL-----DSCYWDHKEAPYSIEDIHDVGLD-----DEEGE 170
DB 418 APTFGRSAYSGLQSTSYLMTTRSFPTYSYSHVOEEOIEIEETIEAKAGEAKAAPAGE 477
QY 171 HEEYDEDAQVVEDEGESEEEEDVSGGDEDEEGYNDGEVDGEEDE---BELGEE 227
DB 478 EEKEEGEEEGAGGEAESEEGEKEGEEEGEETAAEDGESQETAETGEE 537
QY 228 ERGQK 232
DB 538 EKEEK 542

RESULT 9
E86336
hypothetical protein F14010.11 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: E86336
R:theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
ansan, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
```

A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E86336  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-409 <STO>  
A:Cross-references: GB:AE005172; NID:g9558597; PIDN:AAF88160.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 14.5%; Score 176; DB 2; Length 409;  
Best Local Similarity 25.4%; Pred. No. 0.00085;  
Matches 68; Conservative 41; Mismatches 91; Indels 69; Gaps 7;  
QY 2 EMGRRTHSELNRNAPSVDVKELADNSRNEGKLEALTDFFEEFLSKINGLTIS--- 58  
Db 129 EQARLIPSSFHQVDP-DKMLLLAADKINR-----LTLISGFKETISVYT 172  
QY 59 --DLPLKLRKLELRVSGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKQLENKSLDL 116  
Db 173 LVGNPKWTFRK-----NENPKWALDKRIKNVSMGSMFLWNVVAYDG 215  
QY 117 FNCEVTNLDYGE-----NVFKLLQLTYL---DSCVW----- 146  
Db 216 KRLVREMKDILIEGVVHYDMEANSRWVLCSTKWNPSPDSDFYKTPSMVHVGECTIVOA 275  
QY 147 DHKEAPYSDIEDHVEGLDDEEHEEYDEDAQVVEDEEHEEYDEEHEEYDEEYDEEYDEE 206  
Db 276 EEEEEEEEEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEE 335  
QY 207 EGYNDGVDGDEDEELGEERGGOKR 234  
Db 336 EEEEEEEEEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEE 363

## RESULT 10

B44841  
low molecular weight neurofilament protein XNF-L - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 31-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 12-Apr-1995  
C:Accession: B44841  
R:Charnas, L.R.; Szaro, B.G.; Gainer, H.  
J. Neurosci. 12, 3010-3024, 1992  
A:Title: Identification and developmental expression of a novel low molecular weight neurofilament protein XNF-L from the African clawed frog  
A:Reference number: A44841; MUID:92356194; PMID:149494  
A:Accession: B44841  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-544 <CHA>  
A:Experimental source: brain  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBI:110225, NCBI:110226)  
C:Superfamily: cytoskeletal keratin

Query Match 14.4%; Score 174.5; DB 2; Length 544;  
Best Local Similarity 27.6%; Pred. No. 0.0014;  
Matches 72; Conservative 34; Mismatches 96; Indels 59; Gaps 11;  
QY 13 NRAPSVDKELADNSRNEGKLEALTDFFEEFLSKINGLTISDLPKLRKLELRV 72  
Db 302 SRRMLSAKLETEACK--GVNEALQRIQLE--DKQSGETAGQD---AINKLEEL 352  
QY 73 SGGLEVLAEKCPNLTHLYLSGNKIKDLSTIEPLKQLENKSLDLFNCVNTLN----- 125  
Db 353 RNTKSEMAR-----YL--KEYQDLLNVKMLDIEAAYRKLLEGSETRLSFSGVGI 402  
QY 126 -----DYGENFKLLQLTYLDS-----CYWDHKEAPYSDIEDHVEGLDDEE-EGE 170  
Db 403 TSGYTQSPVFGRSAYS-LQSSYMTSRAFTYTSVSHVQEQQLDIETIESRAEAKAE 461  
QY 171 HEEYDEDAQVVEDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEE 213  
Db 462 APEEEEEEAAEEGGEGEABEGEGEAKAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 521

QY 214 EVDGEDEEELGEERGGOKR 234  
Db 522 EEEGESKGDAAEESEKKEK 542

## RESULT 11

T42963  
hypothetical protein 48 - ateline herpesvirus 3 (strain 73)  
C:Species: ateline herpesvirus 3  
A:Variety: strain 73  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
C:Accession: T42963  
R:Albrecht, J.C.; Fleckenstein, B.  
Submitted to the EMBL Data Library, August 1998  
A:Description: Primary structure of the herpesvirus ateles genome.  
A:Reference number: 22274  
A:Accession: T42963  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-792 <ALB>  
A:Cross-references: EMBL:AF083424; PIDN:AAC95573.1  
A:Experimental source: strain 73

Query Match 14.0%; Score 170.5; DB 2; Length 792;  
Best Local Similarity 28.7%; Pred. No. 0.0038;  
Matches 60; Conservative 37; Mismatches 69; Indels 43; Gaps 7;  
QY 26 NSRNECKLALTDFFEEFLSKINGLTISDLPKLRKLELRVSGGLEVLAEKCPN 85  
Db 378 SGRNNKYKGDGANDKKSID-KNESEGG--DHSEINREKNRK-----RKKPN 421  
QY 86 LTHLYLSGNKI--KDLSTIEPLKQLENKSLDLFNCVNTLNNDYGENVFKLLQLTYLDS 143  
Db 422 -----GFRVGDKEVGEKSVKSGGKKS-EKDSDEAEADKDEEN----- 460  
QY 144 CYWDHKEAPYSDIEDHVEGLDDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEE 203  
Db 461 ----KKKGDE 516  
QY 204 EDEEGYNDGVDGDEDEELGEERGGOKR 232  
Db 517 EDE 545

## RESULT 12

A40437  
glutamic acid-rich protein, retinal - bovine  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 14-Feb-1992 #sequence\_revision 14-Feb-1992 #text\_change 05-Nov-1999  
C:Accession: A40437  
R:Sugimoto, Y.; Yatsunami, K.; Tsujimoto, M.; Khorana, H.G.; Ichikawa, A.  
Proc. Natl. Acad. Sci. U.S.A. 88, 3116-3119, 1991  
A:Title: The amino acid sequence of a glutamic acid-rich protein from bovine retina  
A:Reference number: A40437; MUID:91195303; PMID:2014230  
A:Accession: A40437  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-590 <SUG>  
A:Cross-references: GB:M61185; NID:g163077; PIDN:AAA30536.1; PID:g163078

Query Match 13.8%; Score 167.5; DB 2; Length 590;  
Best Local Similarity 43.5%; Pred. No. 0.0042;  
Matches 40; Conservative 18; Mismatches 29; Indels 5; Gaps 3;  
QY 147 DHKEAPYSDIEDHVEGLDDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEE 205  
Db 359 DEEEEDGDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEEHEEYDEE 417  
QY 206 DEEGYNDGVDGDEDEELGEERGGOKR 234  
Db 418 DEEGKEKEEGRKEEG 449

```

Db      316 KTLIEACR--GMNEALEKQLQLELE--DKQNADISAMQD-----TINKLENLNRSTKSEM 366
      80 AEKCPNWLTHLYLSGNKIRKDLSTIEPLKOLENLKSLDLFNCVETNIN----- 125
Db      367 AR-----YL--KEYODLLNVKMLADIEIAAYARKLLEGEETRLSFTSVGTSITSGYSQS 416
      126 --DYGENVFKLLIQLTYLIDSC-----YWDHKEAPYSDIEDHVEGLDDEE-----EGE 170
Db      417 SQVFGSAVSGLOSSLSARSFPAYTSHVQEQTVEETIEATKAEAKDEPPSEGE 476
      171 HEEYDEDAQVVEDEEGEEEBEEDVSGGDEEDGYNDGEVGDGEEDEE-----L 224
Db      477 AEBE-EKEEKEGEEEGAEAEBAKDE--SEDTKEEEGEGGEGEEDTKESEEEEKKEESA 533
      225 GEERQOKR 233
      534 GEEQVAKK 542

RESULT 14
S29795
hypothetical protein 2280 - evening primrose (Oenothera laticarpa subsp. picensis)
C:Species: chloroplast Oenothera laticarpa subsp. picensis (evening primrose)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: S29795; S19983
R:Nimzyk, R.; Schoendorf, T.; Hachtel, W.
Curr. Genet. 23, 265-270, 1993
A:Title: In-frame length mutations associated with short tandem repeats are lo
A:Reference number: S29795; MUID:93169690; PMID:8435856
A:Accession: S29795
A:Molecule type: DNA
A:Residues: 1-721 <NIM>
A:Cross-references: EMBL:X64616; NID:gl4334; PID:gl4335
C:Genetics:
A:Genome: chloroplast
C:Keywords: chloroplast

Query Match 13.6%; Score 165; DB 2: Length 721;
Best Local Similarity 48.2%; Pred. No. 0.0074;
Matches 41; Conservative 9; Mismatches 25; Indels 10; Gaps 3

Qy 154 SDIEDHVEGLDDEEGEHE--EYDEDAQVVEDE--EGEEEEEDEEGEEDVSG-----GD 203
Db 180 SPTEERVEGTGEEVSGTEEEVEGTGTEEGTEGTEDEEVEGTEEGTEEGTEEGTEEGTEEGT 239
Qy 204 EEDEEGYNDGEVGDGEEDEELGEE 228
Db 240 EEEVEGTGTEEGTEEGTEEGTEEGTEEGTEEGTEEGTEEGTEEGTEEGTEEGTEEGTEEG 264

RESULT 15
A36811
hypothetical protein ORF48 - saimirine herpesvirus 1 (strain 11)
C:Species: saimirine herpesvirus 1
A:Note: host Saimiri sciureus (common squirrel monkey)
C>Date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change 08-Oct-1999
C:Accession: A36811
R:Albrecht, J.
submitted to the EMBL Data Library, January 1992
A:Description: Primary structure of the herpesvirus saimirine genome.
A:Reference number: A36806
A:Accession: A36811
A:Molecule type: DNA
A:Residues: 1-797 <ALB>
A:Cross-references: GB:X64346; NID:g60320; PIDN:CAA45671.1; PID:g60369
R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman
J. Virol. 66, 5047-5058, 1992
A:Title: Primary structure of the herpesvirus saimirine genome.
A:Reference number: A37309; MUID:94333688; PMID:1321287
A:Contents: annotation; protein-coding frames
A:Note: neither protein nor nucleotide sequence is given
C:Genetics:

```

A;Gene: 48

```
Query Match      13.6%; Score 165; DB 2; Length 797;
Best Local Similarity 28.8%; Pred. No. 0.0082;
Matches 72; Conservative 37; Mismatches 89; Indels 52; Gaps 13;

QY 22 LAL--DNSRSNGKLEALTDFFEEFLSKINGGLTSTSDL-----PKLKLRLKE 69
   ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 239 LALHPNDKASYSNLIKFLTSNSHREHVTQKVNRAFMQSSLYKIIDTEKNPSPKTKMLM 298
   :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |

QY 70 LAVSG-----GLEVLAE----KCPNLTHLYLSGNKIKDLSTIEPL--KOLENLKSIDLPLNC 119
   :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 299 ISILSGRGIGMDLFCQSQSVLKAPLIDH-----KLSVPSEYEDFDEDEVELCISDDEYDS 352
   :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |

QY 120 EVTN---LNDYGENVFKLL-OLTYLDCYWDHKEAPYSDIEDHVEGLDDDEEGEHE--- 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 353 EDGNLCVLDDSESVNSVALRQVLTVDK---QANEKEYKKIIDKSDDDRRDRDKDEYELEN 409
   ||| ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |

QY 173 EY--DEDAQVVEDEGE-EEEEEGEEDVSGGDEEDEEGYNDG-----EVDGE 218
   ||| ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |
Db 410 EYNRDEEEDGEDEGEDEKEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGE 469
   ||| ||| :| | :| | :| | :| | :| | :| | :| | :| | :| | :| | :| |

QY 219 EDEEELGEE 228
Db 470 DDEDEGEDE 479
```

Search completed: July 7, 2003, 14:58:59  
Job time : 50 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2003, 14:50:37 ; Search time 22 seconds  
(without alignments)  
441.157 Million cell updates/sec

Title: US-09-591-500A-4

Perfect score: 1216

Sequence: 1 MEMGRIHSELNRAPSDVK.....VDGEDEELGEERQKPK 234

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match % | Length | ID           | Description        |
|------------|-------|---------------|--------|--------------|--------------------|
| 1          | 1001  | 82.3          | 249    | 1 PHAL_HUMAN | P39687 homo sapien |
| 2          | 883   | 72.6          | 247    | 1 LANP_RAT   | P49911 rattus norv |
| 3          | 624.5 | 51.4          | 173    | 1 LANP_BOVIN | P51122 bos taurus  |
| 4          | 177.5 | 14.6          | 555    | 1 NFL_CORJA  | Q02916 coturnix co |
| 5          | 174.5 | 14.4          | 544    | 1 NFL_XENLA  | P35616 xenopus lae |
| 6          | 174   | 14.3          | 541    | 1 NFL_RAT    | P19527 rattus norv |
| 7          | 167.5 | 13.8          | 1394   | 1 CNG4_BOVIN | Q28181 bos taurus  |
| 8          | 166.5 | 13.7          | 542    | 1 NFL_MOUSE  | P08551 mus musculu |
| 9          | 165   | 13.6          | 721    | 1 YCP2_OENPI | P31568 oenothera p |
| 10         | 165   | 13.6          | 737    | 1 VG48_HSVSA | Q01033 herpesvirus |
| 11         | 163.5 | 13.4          | 543    | 1 NFL_HUMAN  | P07196 homo sapien |
| 12         | 163   | 13.4          | 678    | 1 GARP_PLAFA | P13816 plasmodium  |
| 13         | 162.5 | 13.4          | 364    | 1 IE68_PVKKA | P24827 pseudorabie |
| 14         | 160   | 13.2          | 587    | 1 RGPI_HUMAN | P46060 homo sapien |
| 15         | 158   | 13.0          | 279    | 1 ASFI_YEAST | P32447 saccharomyc |
| 16         | 157   | 12.9          | 630    | 1 YCF2_OENVI | P31569 oenothera v |
| 17         | 156   | 12.8          | 599    | 1 CENB_MOUSE | P27790 mus musculu |
| 18         | 155   | 12.7          | 599    | 1 CENB_HUMAN | P07199 homo sapien |
| 19         | 155   | 12.7          | 857    | 1 NFL_PIG    | P26053 gallus gall |
| 20         | 154.5 | 12.7          | 548    | 1 NFL_CHICK  | P02547 sus scrofa  |
| 21         | 154.5 | 12.7          | 580    | 1 NFL_XENLA  | O33066 xenopus lae |
| 22         | 154   | 12.7          | 589    | 1 RGPI_MOUSE | P46061 mus musculu |
| 23         | 153   | 12.6          | 539    | 1 Z173_HUMAN | Q12899 homo sapien |
| 24         | 151   | 12.4          | 915    | 1 NFM_HUMAN  | P07197 homo sapien |
| 25         | 150   | 12.3          | 606    | 1 CENB_CRIGR | P48988 cricetus    |
| 26         | 149.5 | 12.3          | 407    | 1 IE68_HSVSA | Q01042 herpesvirus |
| 27         | 149   | 12.3          | 239    | 1 CENB_SHEEP | P49451 ovis aries  |
| 28         | 148.5 | 12.2          | 1220   | 1 IF2P_HUMAN | O60841 homo sapien |
| 29         | 147   | 12.1          | 554    | 1 NFL_BOVIN  | P02548 bos taurus  |
| 30         | 147   | 12.1          | 845    | 1 NFM_RAT    | P12839 rattus norv |
| 31         | 146.5 | 12.0          | 848    | 1 NFM_MOUSE  | P08553 mus musculu |
| 32         | 146   | 12.0          | 743    | 1 ABRA_PLAFA | P22620 plasmodium  |
| 33         | 145.5 | 12.0          | 810    | 1 NFM_BOVIN  | O77788 bos taurus  |

|    |       |      |      |               |                    |
|----|-------|------|------|---------------|--------------------|
| 34 | 145.5 | 12.0 | 3135 | 1 S230_PLAFO  | Q08372 plasmodium  |
| 35 | 144.5 | 11.9 | 1035 | 1 CC68_YEAST  | P32558 saccharomyc |
| 36 | 144   | 11.8 | 593  | 1 SANT_PLAFA7 | Q03400 plasmodium  |
| 37 | 144   | 11.8 | 802  | 1 NAB3_YEAST  | P38996 saccharomyc |
| 38 | 143   | 11.8 | 706  | 1 NUC1_MOUSE  | P09405 mus musculu |
| 39 | 142.5 | 11.7 | 411  | 1 MP62_LYTPI  | P47245 lytechinus  |
| 40 | 142.5 | 11.7 | 1161 | 1 NRDC_RAT    | P47245 rattus norv |
| 41 | 140.5 | 11.6 | 713  | 1 NUC1_MESAU  | P08199 mesocricetu |
| 42 | 140   | 11.5 | 706  | 1 NUC1_HUMAN  | P19338 homo sapien |
| 43 | 140   | 11.5 | 1489 | 1 YGPO_YEAST  | P53115 saccharomyc |
| 44 | 139.5 | 11.5 | 392  | 1 NPT1_YEAST  | P33441 saccharomyc |
| 45 | 138.5 | 11.4 | 712  | 1 NUC1_RAT    | P13383 rattus norv |

## ALIGNMENTS

## RESULT 1

| ID | PHAL_HUMAN  | STANDARD | PRT | 249 AA |
|----|---|----------|-----|--------|
| AC | P39687  |          |     |        |
| DT | 01-FEB-1995 (Rel. 31, Created)  |          |     |        |
| DT | 01-FEB-1995 (Rel. 31, Last sequence update)   |          |     |        |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update)   |          |     |        |
| DE | Potent heat-stable protein phosphatase 2A inhibitor I1PP2A (HLA-DR associated protein I) (PHAPI) (Acidic nuclear phosphoprotein PP32) (Cerebellar leucine rich acidic nuclear protein).   |          |     |        |
| GN | PHAPI OR LANP   |          |     |        |
| OS | Homo sapiens (Human)  |          |     |        |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |          |     |        |
| OX | NCBI-TaxID=9606;  |          |     |        |
| RN | [1]   |          |     |        |
| RP | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.   |          |     |        |
| RX | MEDLINE=94250340; PubMed=8192856;   |          |     |        |
| RA | Vaesens M., Barnikol-Watanabe S., Goetz H., Adil Awmi L., Cole T., Zimmermann B., Kratzin H.D., Hilschmann N.;  |          |     |        |
| RT | "Purification and characterization of two putative HLA class II associated proteins: PHAPI and PHAPII."   |          |     |        |
| RL | Biol. Chem. Hoppe-Seyler 375:113-126(1994).   |          |     |        |
| RN | [2]   |          |     |        |
| RP | SEQUENCE FROM N.A.  |          |     |        |
| RX | TISSUE=Kidney;  |          |     |        |
| RA | MEDLINE=96240314; PubMed=8679524;   |          |     |        |
| RT | Li M., Makkinje A., Damuni Z.;  |          |     |        |
| RL | "Molecular identification of I1PP2A, a novel potent heat-stable inhibitor protein of protein phosphatase 2A.";  |          |     |        |
| RN | Biochemistry 35:6998-7002(1996).  |          |     |        |
| RN | [3]   |          |     |        |
| RP | SEQUENCE FROM N.A.  |          |     |        |
| RA | Chen T.H., Brody J.R., Romantsev F.E., Yu J.G., Kayler A.E., Voneiff E., Kuhnada F.P., Pasternack G.R.;   |          |     |        |
| RL | Submitted (JAN-1987) to the EMBL/GenBank/DBJ databases.   |          |     |        |
| RN | [4]   |          |     |        |
| RP | SEQUENCE FROM N.A.  |          |     |        |
| RX | MEDLINE=98013170; PubMed=9353121;   |          |     |        |
| RA | Matilla A., Koshy B.T., Cummings C.J., Isobe T., Orr H.T., Zoghbi H.Y.;   |          |     |        |
| RL | "The cerebellar leucine-rich acidic nuclear protein interacts with ataxin-1.";  |          |     |        |
| RN | Nature 389:974-978(1997).   |          |     |        |
| CC | -!- FUNCTION: INHIBITOR PROTEIN OF PROTEIN PHOSPHATASE 2A.  |          |     |        |
| CC | -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC, DIFFUSELY DISTRIBUTED IN THE CYTOSOL.  |          |     |        |
| CC | -!- PTM: PHOSPHORYLATED ON SERINE RESIDUES.   |          |     |        |
| CC | -!- PTM: THE N-TERMINUS IS BLOCKED.   |          |     |        |
| CC | -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).  |          |     |        |
| CC | -!- SIMILARITY: HOMOLOG OF RAT LANP.  |          |     |        |
| CC | -----   |          |     |        |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way |          |     |        |

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X75090; CAA52981.1; -;  
 CC EMBL; U60823; AAC50570.1; -;  
 CC EMBL; U73477; AAB39706.1; -;  
 CC EMBL; AF025684; AAB91548.1; -;  
 CC PIR; S43309; S43309.  
 CC PIR; S37222; S37222.  
 CC Genew; HGNC:13233; ANP32A.  
 CC MIM; 600832; -;  
 CC InterPro; IPR001611; LRR.  
 CC InterPro; IPR003603; LRRcap.  
 CC Pfam; PF00560; LRR; 2.  
 CC SMART; SM00446; LRRcap; 1.  
 CC Phosphorylation; Leucine-rich repeat; Repeat; Nuclear protein.  
 CC REPEAT 44 63 LRR 1.  
 CC REPEAT 64 86 LRR 2.  
 CC REPEAT 87 111 LRR 3.  
 CC REPEAT 115 138 LRR 4.  
 CC DOMAIN 168 249 ASP/GLU-RICH (HIGHLY ACIDIC).  
 CC SEQUENCE 249 AA; 28585 MW; CA2D1A756FBAEA04 CRC64;

Query Match 82.3%; Score 1001; DB 1; Length 249;  
 Best Local Similarity 86.1%; Pred. No. 1.4e-54;  
 Matches 205; Conservative 7; Mismatches 22; Indels 4; Gaps 2;  
 QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNECKLEALTDDEFEELFLSKINGLTSIDL 60  
 DB 1 MEMGRIHSELNRAPSDVKELALDNRSGNECKLEALTDDEFEELFLSKINGLTSIDL 60  
 QY 61 PKI-KLRKLEL---RVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKLSIDL 116  
 DB 61 PKLNLKLELSDNRSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKLSIDL 120  
 QY 117 FNCVNTLNDYGNVFKLLQLTYLDSCYWDHKEAPYSDIEDHVGLEDDEEGEREYD 176  
 DB 121 FNCVNTLNDYGNVFKLLPQTYLDGYDRDKAPSDAEGYVGLDDEEGEREYD 180  
 QY 177 EDAQVDEDEEGEREYDVGDEDEGYNDGVDGDEDEEELGEEERGOKR 234  
 DB 181 EDAQVDEDEEGEREYDVGDEDEGYNDGVDGDEDEEELGEEERGOKR 238

RESULT 2  
 LAMP\_RAT LAMP\_RAT STANDARD; PRT; 247 AA.  
 ID LAMP\_RAT  
 AC P49911;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Leucine-rich acidic nuclear protein.  
 GN LAMP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Brain;  
 RX MEDLINE=95024022; PubMed=7937870;  
 RA Matsuo K., Taoka M., Satozawa N., Nakayama H., Ichimura T.,  
 RA Takahashi N., Yamakuni T., Song S.-Y., Isobe T.;  
 RT "A nuclear factor containing the leucine-rich repeats expressed in  
 RT murine cerebellar neurons";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9670-9674(1994).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN A SIGNAL TRANSDUCTION PATHWAY THAT  
 CC DIRECTS DIFFERENTIATION OF CEREBELLAR NEURONS.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: WIDELY DISTRIBUTED IN THE CENTRAL NERVOUS  
 CC SYSTEM, WITH AN ABUNDANT EXPRESSION IN THE CEREBELLUM.  
 CC -!- DEVELOPMENTAL STAGE: IT IS MODERATELY EXPRESSED IN THE CEREBELLUM

CC ON POSTNATAL DAY 7 IN THE EXTERNAL GRANULE AND PERKINJE CELLS,  
 CC INCREASES IN THE SECOND POSTNATAL WEEK AND THEREAFTER DECREASES  
 CC TO AN ADULT LEVEL.  
 CC -!- SIMILARITY: CONTAINS 4 LEUCINE-RICH REPEATS (LRR).  
 CC -!- SIMILARITY: HOMOLOG OF HUMAN PHAP-1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D32209; BAA06908.1; -;  
 CC InterPro; IPR001611; LRR.  
 CC InterPro; IPR003603; LRRcap.  
 CC Pfam; PF00560; LRR; 2.  
 CC SMART; SM00446; LRRcap; 1.  
 CC Phosphorylation; Leucine-rich repeat; Repeat; Nuclear protein.  
 CC REPEAT 44 63 LRR 1.  
 CC REPEAT 64 86 LRR 2.  
 CC REPEAT 87 111 LRR 3.  
 CC REPEAT 115 138 LRR 4.  
 CC DOMAIN 165 247 ASP/GLU-RICH (HIGHLY ACIDIC).  
 CC SEQUENCE 247 AA; 28564 MW; 0919739645557462 CRC64;

Query Match 72.6%; Score 883; DB 1; Length 247;  
 Best Local Similarity 77.3%; Pred. No. 2e-47;  
 Matches 184; Conservative 16; Mismatches 32; Indels 6; Gaps 3;  
 QY 1 MEMGRIHSELNRAPSDVKELALDNRSGNECKLEALTDDEFEELFLSKINGLTSIDL 60  
 DB 1 MEMGRIHSELNRAPSDVKELALDNRSGNECKLEALTDDEFEELFLSKINGLTSIDL 60  
 QY 61 PKI-KLRKLEL---RVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKLSIDL 116  
 DB 61 PKLNLKLELSDNRSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKLENKLSIDL 120  
 QY 117 FNCVNTLNDYGNVFKLLQLTYLDSCYWDHKEAPYSDIEDHVGLEDDEEGEREYD 176  
 DB 121 FNCVNTLNDYGNVFKLLPQTYLDGYDRDKAPSDSVEGYVE--DDDEDEDEEYD 178  
 QY 177 EDAQVDEDEEGEREYDVGDEDEGYNDGVDGDEDEEELGEEERGOKR 234  
 DB 179 EYAQVDEDEEGEREYDVGDEDEGYNDGVDGDEDEEELGEEERGOKR 236

RESULT 3  
 LAMP\_BOVIN LAMP\_BOVIN STANDARD; PRT; 173 AA.  
 ID LAMP\_BOVIN  
 AC P51122;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Leucine-rich acidic nuclear protein (Fragments).  
 GN LAMP.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=95024022; PubMed=7937870;  
 RA Matsuo K., Taoka M., Satozawa N., Nakayama H., Ichimura T.,  
 RA Takahashi N., Yamakuni T., Song S.-Y., Isobe T.;  
 RT "A nuclear factor containing the leucine-rich repeats expressed in  
 RT murine cerebellar neurons";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9670-9674(1994).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN A SIGNAL TRANSDUCTION PATHWAY THAT  
 CC DIRECTS DIFFERENTIATION OF CEREBELLAR NEURONS.





|    |  |  |
|----|--|--|
| RT | weight neuronal intermediate filament protein expressed in xenopus laevis."  |  |
| RL | J. Neurosci. 12:3010-3024(1992).   |  |
| CC | -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M, AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.   |  |
| CC | -1- MISCELLANEOUS: NF-L IS THE MOST ABUNDANT OF THE THREE NEUROFILAMENT PROTEINS AND, AS THE OTHER NONEPITHELIAL INTERMEDIATE FILAMENT PROTEINS, IT CAN FORM HOMOPOLYMERIC 10-NM FILAMENTS.  |  |
| CC | -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.   |  |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ). |  |
| DR | EMBL; M86654; AAA03018.1; -.   |  |
| DR | PIR; B44841; B44841.   |  |
| DR | InterPro; IPR001664; IF.   |  |
| DR | Pfam; PF00038; filament; 1.  |  |
| DR | PROSITE; PS00226; IF; 1.   |  |
| KW | Intermediate filament; Coiled coil; Neurone.   |  |
| FT | DOMAIN 1 87  |  |
| FT | DOMAIN 88 390  |  |
| FT | DOMAIN 391 544   |  |
| FT | DOMAIN 545 725   |  |
| FT | DOMAIN 726 877   |  |
| FT | DOMAIN 878 1029  |  |
| FT | DOMAIN 1030 1181   |  |
| FT | DOMAIN 1182 1333   |  |
| FT | DOMAIN 1334 1485   |  |
| FT | DOMAIN 1486 1637   |  |
| FT | DOMAIN 1638 1789   |  |
| FT | DOMAIN 1790 1941   |  |
| FT | DOMAIN 1942 2093   |  |
| FT | DOMAIN 2094 2245   |  |
| FT | DOMAIN 2246 2397   |  |
| FT | DOMAIN 2398 2549   |  |
| FT | DOMAIN 2550 2701   |  |
| FT | DOMAIN 2702 2853   |  |
| FT | DOMAIN 2854 3005   |  |
| FT | DOMAIN 3006 3157   |  |
| FT | DOMAIN 3158 3309   |  |
| FT | DOMAIN 3310 3461   |  |
| FT | DOMAIN 3462 3613   |  |
| FT | DOMAIN 3614 3765   |  |
| FT | DOMAIN 3766 3917   |  |
| FT | DOMAIN 3918 4069   |  |
| FT | DOMAIN 4070 4221   |  |
| FT | DOMAIN 4222 4373   |  |
| FT | DOMAIN 4374 4525   |  |
| FT | DOMAIN 4526 4677   |  |
| FT | DOMAIN 4678 4829   |  |
| FT | DOMAIN 4830 4981   |  |
| FT | DOMAIN 4982 5133   |  |
| FT | DOMAIN 5134 5285   |  |
| FT | DOMAIN 5286 5437   |  |
| FT | DOMAIN 5438 5589   |  |
| FT | DOMAIN 5590 5741   |  |
| FT | DOMAIN 5742 5893   |  |
| FT | DOMAIN 5894 6045   |  |
| FT | DOMAIN 6046 6197   |  |
| FT | DOMAIN 6198 6349   |  |
| FT | DOMAIN 6350 6501   |  |
| FT | DOMAIN 6502 6653   |  |
| FT | DOMAIN 6654 6805   |  |
| FT | DOMAIN 6806 6957   |  |
| FT | DOMAIN 6958 7109   |  |
| FT | DOMAIN 7110 7261   |  |
| FT | DOMAIN 7262 7413   |  |
| FT | DOMAIN 7414 7565   |  |
| FT | DOMAIN 7566 7717   |  |
| FT | DOMAIN 7718 7869   |  |
| FT | DOMAIN 7870 8021   |  |
| FT | DOMAIN 8022 8173   |  |
| FT | DOMAIN 8174 8325   |  |
| FT | DOMAIN 8326 8477   |  |
| FT | DOMAIN 8478 8629   |  |
| FT | DOMAIN 8630 8781   |  |
| FT | DOMAIN 8782 8933   |  |
| FT | DOMAIN 8934 9085   |  |
| FT | DOMAIN 9086 9237   |  |
| FT | DOMAIN 9238 9389   |  |
| FT | DOMAIN 9390 9541   |  |
| FT | DOMAIN 9542 9693   |  |
| FT | DOMAIN 9694 9845   |  |
| FT | DOMAIN 9846 9997   |  |
| FT | DOMAIN 10000 10151   |  |
| FT | DOMAIN 10152 10303   |  |
| FT | DOMAIN 10304 10455   |  |
| FT | DOMAIN 10456 10607   |  |
| FT | DOMAIN 10608 10759   |  |
| FT | DOMAIN 10760 10911   |  |
| FT | DOMAIN 10912 11063   |  |
| FT | DOMAIN 11064 11215   |  |
| FT | DOMAIN 11216 11367   |  |
| FT | DOMAIN 11368 11519   |  |
| FT | DOMAIN 11520 11671   |  |
| FT | DOMAIN 11672 11823   |  |
| FT | DOMAIN 11824 11975   |  |
| FT | DOMAIN 11976 12127   |  |
| FT | DOMAIN 12128 12279   |  |
| FT | DOMAIN 12280 12431   |  |
| FT | DOMAIN 12432 12583   |  |
| FT | DOMAIN 12584 12735   |  |
| FT | DOMAIN 12736 12887   |  |
| FT | DOMAIN 12888 13039   |  |
| FT | DOMAIN 13040 13191   |  |
| FT | DOMAIN 13192 13343   |  |
| FT | DOMAIN 13344 13495   |  |
| FT | DOMAIN 13496 13647   |  |
| FT | DOMAIN 13648 13799   |  |
| FT | DOMAIN 13800 13951   |  |
| FT | DOMAIN 13952 14103   |  |
| FT | DOMAIN 14104 14255   |  |
| FT | DOMAIN 14256 14407   |  |
| FT | DOMAIN 14408 14559   |  |
| FT | DOMAIN 14560 14711   |  |
| FT | DOMAIN 14712 14863   |  |
| FT | DOMAIN 14864 15015   |  |
| FT | DOMAIN 15016 15167   |  |
| FT | DOMAIN 15168 15319   |  |
| FT | DOMAIN 15320 15471   |  |
| FT | DOMAIN 15472 15623   |  |
| FT | DOMAIN 15624 15775   |  |
| FT | DOMAIN 15776 15927   |  |
| FT | DOMAIN 15928 16079   |  |
| FT | DOMAIN 16080 16231   |  |
| FT | DOMAIN 16232 16383   |  |
| FT | DOMAIN 16384 16535   |  |
| FT | DOMAIN 16536 16687   |  |
| FT | DOMAIN 16688 16839   |  |
| FT | DOMAIN 16840 16991   |  |
| FT | DOMAIN 16992 17143   |  |
| FT | DOMAIN 17144 17295   |  |
| FT | DOMAIN 17296 17447   |  |
| FT | DOMAIN 17448 17599   |  |
| FT | DOMAIN 17600 17751   |  |
| FT | DOMAIN 17752 17903   |  |
| FT | DOMAIN 17904 18055   |  |
| FT | DOMAIN 18056 18207   |  |
| FT | DOMAIN 18208 18359   |  |
| FT | DOMAIN 18360 18511   |  |
| FT | DOMAIN 18512 18663   |  |
| FT | DOMAIN 18664 18815   |  |
| FT | DOMAIN 18816 18967   |  |
| FT | DOMAIN 18968 19119   |  |
| FT | DOMAIN 19120 19271   |  |
| FT | DOMAIN 19272 19423   |  |
| FT | DOMAIN 19424 19575   |  |
| FT | DOMAIN 19576 19727   |  |
| FT | DOMAIN 19728 19879   |  |
| FT | DOMAIN 19880 20031   |  |
| FT | DOMAIN 20032 20183   |  |
| FT | DOMAIN 20184 20335   |  |
| FT | DOMAIN 20336 20487   |  |
| FT | DOMAIN 20488 20639   |  |
| FT | DOMAIN 20640 20791   |  |
| FT | DOMAIN 20792 20943   |  |
| FT | DOMAIN 20944 21095   |  |
| FT | DOMAIN 21096 21247   |  |
| FT | DOMAIN 21248 21399   |  |
| FT | DOMAIN 21400 21551   |  |
| FT | DOMAIN 21552 21703   |  |
| FT | DOMAIN 21704 21855   |  |
| FT | DOMAIN 21856 22007   |  |
| FT | DOMAIN 22008 22159   |  |
| FT | DOMAIN 22160 22311   |  |
| FT | DOMAIN 22312 22463   |  |
| FT | DOMAIN 22464 22615   |  |
| FT | DOMAIN 22616 22767   |  |
| FT | DOMAIN 22768 22919   |  |
| FT | DOMAIN 22920 23071   |  |
| FT | DOMAIN 23072 23223   |  |
| FT | DOMAIN 23224 23375   |  |
| FT | DOMAIN 23376 23527   |  |
| FT | DOMAIN 23528 23679   |  |
| FT | DOMAIN 23680 23831   |  |
| FT | DOMAIN 23832 23983   |  |
| FT | DOMAIN 23984 24135   |  |
| FT | DOMAIN 24136 24287   |  |
| FT | DOMAIN 24288 24439   |  |
| FT | DOMAIN 24440 24591   |  |
| FT | DOMAIN 24592 24743   |  |
| FT | DOMAIN 24744 24895   |  |
| FT | DOMAIN 24896 25047   |  |
| FT | DOMAIN 25048 25199   |  |
| FT | DOMAIN 25200 25351   |  |
| FT | DOMAIN 25352 25503   |  |
| FT | DOMAIN 25504 25655   |  |
| FT | DOMAIN 25656 25807   |  |
| FT | DOMAIN 25808 25959   |  |
| FT | DOMAIN 25960 26111   |  |
| FT | DOMAIN 26112 26263   |  |
| FT | DOMAIN 26264 26415   |  |
| FT | DOMAIN 26416 26567   |  |
| FT | DOMAIN 26568 26719   |  |
| FT | DOMAIN 26720 26871   |  |
| FT | DOMAIN 26872 27023   |  |
| FT | DOMAIN 27024 27175   |  |
| FT | DOMAIN 27176 27327   |  |
| FT | DOMAIN 27328 27479   |  |
| FT | DOMAIN 27480 27631   |  |
| FT | DOMAIN 27632 27783   |  |
| FT | DOMAIN 27784 27935   |  |
| FT | DOMAIN 27936 28087   |  |
| FT | DOMAIN 28088 28239   |  |
| FT | DOMAIN 28240 28391   |  |
| FT | DOMAIN 28392 28543   |  |
| FT | DOMAIN 28544 28695   |  |
| FT | DOMAIN 28696 28847   |  |
| FT | DOMAIN 28848 28999   |  |
| FT | DOMAIN 29000 29151   |  |
| FT | DOMAIN 29152 29303   |  |
| FT | DOMAIN 29304 29455   |  |
| FT | DOMAIN 29456 29607   |  |
| FT | DOMAIN 29608 29759   |  |
| FT | DOMAIN 29760 29911   |  |
| FT | DOMAIN 29912 30063   |  |
| FT | DOMAIN 30064 30215   |  |
| FT | DOMAIN 30216 30367   |  |
| FT | DOMAIN 30368 30519   |  |
| FT | DOMAIN 30520 30671   |  |
| FT | DOMAIN 30672 30823   |  |
| FT | DOMAIN 30824 30975   |  |
| FT | DOMAIN 30976 31127   |  |
| FT | DOMAIN 31128 31279   |  |
| FT | DOMAIN 31280 31431   |  |
| FT | DOMAIN 31432 31583   |  |
| FT | DOMAIN 31584 31735   |  |
| FT | DOMAIN 31736 31887   |  |
| FT | DOMAIN 31888 32039   |  |
| FT | DOMAIN 32040 32191   |  |
| FT | DOMAIN 32192 32343   |  |
| FT | DOMAIN 32344 32495   |  |
| FT | DOMAIN 32496 32647   |  |
| FT | DOMAIN 32648 32799   |  |
| FT | DOMAIN 32800 32951   |  |
| FT | DOMAIN 32952 33103   |  |
| FT | DOMAIN 33104 33255   |  |
| FT | DOMAIN 33256 33407   |  |
| FT | DOMAIN 33408 33559   |  |
| FT | DOMAIN 33560 33711   |  |
| FT | DOMAIN 33712 33863   |  |
| FT | DOMAIN 33864 34015   |  |
| FT | DOMAIN 34016 34167   |  |
| FT | DOMAIN 34168 34319   |  |
| FT | DOMAIN 34320 34471   |  |
| FT | DOMAIN 34472 34623   |  |
| FT | DOMAIN 34624 34775   |  |
| FT | DOMAIN 34776 34927   |  |
| FT | DOMAIN 34928 35079   |  |
| FT | DOMAIN 35080 35231   |  |
| FT | DOMAIN 35232 35383   |  |
| FT | DOMAIN 35384 35535   |  |
| FT | DOMAIN 35536 35687   |  |
| FT | DOMAIN 35688 35839   |  |
| FT | DOMAIN 35840 35991   |  |
| FT | DOMAIN 35992 36143   |  |
| FT | DOMAIN 36144 36295   |  |
| FT | DOMAIN 36296 36447   |  |
| FT | DOMAIN 36448 36599   |  |
| FT | DOMAIN 36600 36751   |  |
| FT | DOMAIN 36752 36903   |  |
| FT | DOMAIN 36904 37055   |  |
| FT | DOMAIN 37056 37207   |  |
| FT | DOMAIN 37208 37359   |  |
| FT | DOMAIN 37360 37511   |  |
| FT | DOMAIN 37512 37663   |  |
| FT | DOMAIN 37664 37815   |  |
| FT | DOMAIN 37816 37967   |  |
| FT | DOMAIN 37968 38119   |  |
| FT | DOMAIN 38120 38271   |  |
| FT | DOMAIN 38272 38423   |  |
| FT | DOMAIN 38424 38575   |  |
| FT | DOMAIN 38576 38727   |  |
| FT | DOMAIN 38728 38879   |  |
| FT | DOMAIN 38880 39031   |  |
| FT | DOMAIN 39032 39183   |  |
| FT | DOMAIN 39184 39335   |  |
| FT | DOMAIN 39336 39487   |  |
| FT | DOMAIN 39488 39639   |  |
| FT | DOMAIN 39640 39791   |  |
| FT | DOMAIN 39792 39943   |  |
| FT | DOMAIN 39944 40095   |  |
| FT | DOMAIN 40096 40247   |  |
| FT | DOMAIN 40248 40399   |  |
| FT | DOMAIN 40400 40551   |  |
| FT | DOMAIN 40552 40703   |  |
| FT | DOMAIN 40704 40855   |  |
| FT | DOMAIN 40856 41007   |  |
| FT | DOMAIN 41008 41159   |  |
| FT | DOMAIN 41160 41311   |  |
| FT | DOMAIN 41312 41463   |  |
| FT | DOMAIN 41464 41615   |  |
| FT | DOMAIN 41616 41767   |  |
| FT | DOMAIN 41768 41919   |  |
| FT | DOMAIN 41920 42071   |  |
| FT | DOMAIN 42072 42223   |  |
| FT | DOMAIN 42224 42375   |  |
| FT | DOMAIN 42376 42527   |  |
| FT | DOMAIN 42528 42679   |  |
| FT | DOMAIN 42680 42831   |  |
| FT | DOMAIN 42832 42983   |  |
| FT | DOMAIN 42984 43135   |  |
| FT | DOMAIN 43136 43287   |  |
| FT | DOMAIN 43288 43439   |  |
| FT | DOMAIN 43440 43591   |  |
| FT | DOMAIN 43592 43743   |  |
| FT | DOMAIN 43744 43895   |  |
| FT | DOMAIN 43896 44047   |  |
| FT | DOMAIN 44048 44199   |  |
| FT | DOMAIN 44200 44351   |  |
| FT | DOMAIN 44352 44503   |  |
| FT | DOMAIN 44504 44655   |  |
| FT | DOMAIN 44656 44807   |  |
| FT | DOMAIN 44808 44959   |  |
| FT | DOMAIN 44960 45111   |  |
| FT | DOMAIN 45112 45263   |  |
| FT | DOMAIN 45264 45415   |  |
| FT | DOMAIN 45416 45567   |  |
| FT | DOMAIN 45568 45719   |  |
| FT | DOMAIN 45720 45871   |  |
| FT | DOMAIN 45872 46023   |  |
| FT | DOMAIN 46024 46175   |  |
| FT | DOMAIN 46176 46327   |  |
| FT | DOMAIN 46328 46479   |  |
| FT | DOMAIN 46480 46631   |  |
| FT | DOMAIN 46632 46783   |  |
| FT | DOMAIN 46784 46935   |  |
| FT | DOMAIN 46936 47087   |  |
| FT | DOMAIN 47088 47239   |  |
| FT | DOMAIN 47240 47391   |  |
| FT | DOMAIN 47392 47543   |  |
| FT | DOMAIN 47544 47695   |  |
| FT | DOMAIN 47696 47847   |  |
| FT | DOMAIN 47848 47999   |  |
| FT | DOMAIN 48000 48151   |  |
| FT | DOMAIN 48152 48303   |  |
| FT | DOMAIN 48304 48455   |  |
| FT | DOMAIN 48456 48607   |  |
| FT | DOMAIN 48608 48759   |  |
| FT | DOMAIN 48760 48911   |  |
| FT | DOMAIN 48912 49063   |  |
| FT | DOMAIN 49064 49215   |  |
| FT | DOMAIN 49216 49367   |  |
| FT | DOMAIN 49368 49519   |  |
| FT | DOMAIN 49520 49671   |  |
| FT | DOMAIN 49672 49823   |  |
| FT | DOMAIN 49824 49975   |  |
| FT | DOMAIN 49976 50127   |  |
| FT | DOMAIN 50128 50279   |  |
| FT | DOMAIN 50280 50431   |  |
| FT | DOMAIN 50432 50583   |  |
| FT | DOMAIN 50584 50735   |  |
| FT | DOMAIN 50736 50887   |  |
| FT | DOMAIN 50888 51039   |  |
| FT | DOMAIN 51040 51191   |  |
| FT | DOMAIN 51192 51343   |  |
| FT | DOMAIN 51344 51495   |  |
| FT | DOMAIN 51496 51647   |  |
| FT | DOMAIN 51648 51799   |  |
| FT | DOMAIN 51800 51951   |  |
| FT | DOMAIN 51952 52103   |  |
| FT | DOMAIN 52104 52255   |  |
| FT | DOMAIN 52256 52407   |  |
| FT | DOMAIN 52408 52559   |  |
| FT | DOMAIN 52560 52711   |  |
| FT | DOMAIN 52712 52863   |  |
| FT | DOMAIN 52864 53015   |  |
| FT | DOMAIN 53016 53167   |  |
| FT | DOMAIN 53168 53319   |  |
| FT | DOMAIN 53320 53471   |  |
| FT | DOMAIN 53472 53623   |  |
| FT | DOMAIN 53624 53775   |  |
| FT | DOMAIN 53776 53927   |  |
| FT | DOMAIN 53928 54079   |  |
| FT | DOMAIN 54080 54231   |  |
| FT | DOMAIN 54232 54383   |  |
| FT | DOMAIN 54384 54535   |  |
| FT | DOMAIN 54536 54687   |  |
| FT | DOMAIN 54688 54839   |  |
| FT | DOMAIN 54840 54991   |  |
| FT | DOMAIN 54992 55143   |  |
| FT | DOMAIN 55144 55295   |  |
| FT | DOMAIN 55296 55447   |  |
| FT | DOMAIN 55448 55599   |  |
| FT | DOMAIN 55600 55751   |  |
| FT | DOMAIN 55752 55903   |  |
| FT | DOMAIN 55904 56055   |  |
| FT | DOMAIN 56056 56207   |  |
| FT |  |  |

```

FT DOMAIN 93 124 COIL 1A.
FT DOMAIN 125 137 LINKER 1.
FT DOMAIN 138 233 COIL 1B.
FT DOMAIN 234 252 LINKER 12.
FT DOMAIN 253 271 COIL 2A.
FT DOMAIN 272 280 COIL 2.
FT DOMAIN 281 396 COIL 2B.
FT DOMAIN 397 443 TAIL, SUBDOMAIN A.
FT DOMAIN 444 541 TAIL, SUBDOMAIN B (ACIDIC).
FT CARBOHYD 20 20 O-LINKED (GLNAC).
FT CARBOHYD 26 26 /FTID-CAR_000128.
FT CARBOHYD 381 391 O-LINKED (GLNAC).
FT SITE 197 202 EPTIPE (RECOGNIZED BY IF-SPECIFIC
FT CONFLICT 399 399 MONOCLONAL ANTIBODY).
FT CONFLICT 476 476 GADEAA -> KARMSS (IN REF. 2).
FT CONFLICT 476 476 R -> K (IN REF. 2).
FT CONFLICT 480 483 A -> E (IN REF. 2).
SQ SEQUENCE 541 AA: 61204 MW: 0D17839AF226918A CRC64;
                                     14.3% Score 174; DB 1; Length 541;
Query Match Best Local Similarity 26.4% Pred. No. 0.00061;
Matches 65; Conservative 40; Mismatches 89; Indels 52; Gaps 10;
Qy 20 KEALDNRSENGKLEALDTDFEFLFKSLKINGLGLTISDLPLKLRKLELRVSGGLEVL 79
Db 315 KTLIEACR---GMNEALEKQLQLE--DKONADISAMQD---TINKLENELRSTKSEM 365
Qy 80 AERCPNTHLYLSGNKTKDLSTIEPLKQLENKSLDLFNCFVTLN-----125
Db 366 AR-----YL---KEYQDILNVKMLDIEAAYRKLLGEETRLSFTSGVTSIGSYS 415
Qy 126 --DYGENVFKLLQLTYLDSC-----YNDHKEAPYSDIEDHVEGLDDEE-----EGE 170
Db 416 SQVFGSAVSSGLSSVYLSARAFPAYTSHVQEEQSEVEETATKAEDKPPSEGE 475
Qy 171 HEEYDDAOVDEGEEREEEREEEDVSGSGDEDEGYNDGEVDGEED---EEELGEE 227
Db 476 AEEE-EKEEKEEGEEGAEEAAKDESEDAKEEGEGEEDTKSEEEKKEESAGEE 534
Qy 228 ERGOKR 233
Db 535 QAAKK 540

RESULT 7
CNG4_BOVIN STANDARD; PRT: 1394 AA.
ID CNG4_BOVIN STANDARD; Q03861;
AC Q28181; Q28082; Q03861;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 240 kDa protein of rod photoreceptor CNG-channel [Contains: Glutamic
DE acid-rich protein (GARP); Cyclic-nucleotide-gated cation channel 4
DE (CNG channel 4) (CNG-4) (Cyclic nucleotide-gated cation channel
DE modulatory subunit)].
DE CNGBI OR CNGG4.
GN Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=96009859; PubMed=7546742;
RA Koerschen H.G., Illing M., Seifert R., Sesti F., Williams A.,
RA Gotzes S., Colville C., Mueller F., Dose A., Godde M., Molday L.,
RA Kaupp U.B., Molday R.S.;
RA "A 240 kDa protein represents the complete beta subunit of the cyclic
RT nucleotide-gated channel from rod photoreceptor.";
RN Neuron 15:627-636(1995).
[2]

```

```

RP SEQUENCE OF 454-1394 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=96198098; PubMed=8626431;
RA Biel M., Zong X., Ludwig A., Sautter A., Hofmann F.;
RT "Molecular cloning and expression of the modulatory subunit of the
RL cyclic nucleotide-gated cation channel.";
RN J. Biol. Chem. 271:6349-6355(1996).
[3]
RP SEQUENCE OF 1-590 FROM N.A.
RC TISSUE=Retina;
RA Sugimoto Y., Yatsunami K., Tsujimoto M., Khorana H.G., Ichikawa A.;
RL Submitted (XXX-1991) to the EMBL/Genbank/DBJ databases.
CC -!- SUBUNIT: FORMS FUNCTIONAL HETEROOLIGOMERIC CHANNELS WITH CNG3.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; CNG4C (SHOWN HERE), CNG4D
CC AND CNG4E; ARE PRODUCED BY ALTERNATIVE SPLICING. CNG4D IS BY FAR
CC THE MOST FREQUENT FORM (CNG4D:CNG4C:CNG4E = 20:2:1) IN TESTIS.
CC -!- TISSUE SPECIFICITY: RETINA, TESTIS, KIDNEY, HEART, AND BRAIN.
CC -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X89626; CAA61769.1; -;
DR EMBL; X94707; CAA64367.1; -;
DR EMBL; M61185; AAA30536.1; -;
DR InterPro: IPR000636; M+channel_nlg.
DR InterPro: IPR000995; CNGP_binding.
DR Pfam: PF00027; CNGP_binding; 1.
DR Pfam: PF00520; Ion_trans; 1.
DR SMART; SM00100; CNGP; 1.
DR PROSITE; PS00888; CNGP_BINDING_1; 1.
DR PROSITE; PS00889; CNGP_BINDING_2; 1.
DR PROSITE; PS00042; CNGP_BINDING_3; 1.
KW Ionic channel; Ion transport; CAMP-binding; Transmembrane;
KW Multigene family; Alternative splicing.
FT CHAIN 1 590
FT CHAIN 454 1394
FT DOMAIN 1 767
FT TRANSMEM 768 800
FT DOMAIN 787 800
FT TRANSMEM 801 819
FT DOMAIN 821 844
FT TRANSMEM 845 864
FT DOMAIN 865 901
FT TRANSMEM 902 924
FT TRANSMEM 925 968
FT TRANSMEM 969 988
FT DOMAIN 989 1072
FT TRANSMEM 1073 1093
FT DOMAIN 1094 1394
FT NP_BIND 1081 1219
FT BINDING 1141 1141
FT BINDING 1153 1153
FT CARBOHYD 1067 1067
FT VARSPLIC 515 532
FT VARSPLIC 522 530
FT CONFLICT 341 341
FT CONFLICT 454 465
FT CONFLICT 482 482
FT CONFLICT 499 499
FT CONFLICT 572 590
FT CONFLICT 1283 1283
FT CONFLICT 1289 1289
FT CONFLICT 1336 1336
FT CONFLICT 1338 1338

```



```

ID YCF2_OBNPI STANDARD; PRT; 721 AA.
AC P31568;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ycf2 (ORF 2280) (Fragment).
GN YCF2.
OS Oenothera picensis (Oenothera odorata).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Myrtales; Onagraceae; Oenothera.
OX NCBI_TaxID=3946;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93169690; PubMed=8435856;
RA Ninzyk R., Shoendorf T., Hachtel W.;
RT "In-frame length mutations associated with short tandem repeats are
RT located in unassigned open reading frames of Oenothera chloroplast
RT DNA.";
RL Curr. Genet. 23:265-270(1993).
CC -!- FUNCTION: NOT YET KNOWN.
CC -!- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64616; CAA45898.1; .
DR PIR; S29795; S29795.
KW Chloroplast; Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 721 AA; 82900 MW; 279A3FC1AA06F9C9 CRC64;

Query Match 13.6%; Score 165; DB 1; Length 721;
Best Local Similarity 48.2%; Pred. No. 0.0029;
Matches 41; Conservative 9; Mismatches 25; Indels 10; Gaps 3;

Qy 154 SDIEDHVGGLDDEEGEHE--EYDEDAQVVEDE--EGEDEEGEEDVSG-----GD 203
| : | | : | | : | : | : | | | | | | | | | | | | | | | | | |
Db 180 SPTEEEVEGTEEEVEEGTEEEVEEGTEEEVEEGTEEEVEEGTEEEVEEGTEEEVEGT 239
| : | | : | | : | | : | | | | | | | | | | | | | | | | | |
Qy 204 EEEDEGNDVGDDEEDELGEE 228
| : | | : | | : | | | | | | | | | | | | | | | | | |
Db 240 EEEVEGTEDEEGTEDEEGTEEE 264
| : | | : | | : | | | | | | | | | | | | | | | | | |

RESULT 10
VG48_HSVSA STANDARD; PRT; 797 AA.
ID VG48_HSVSA
AC Q01033;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Hypothetical gene 48 protein.
GN 48 OR EDL5.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
CC -!- SIMILARITY: TO EBV BRFF2.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64346; CAA45671.1; .
DR PIR; A36811; A36811.
KW Hypothetical protein.
SQ SEQUENCE 797 AA; 88867 MW; 9E294234AD850E23 CRC64;

Query Match 13.6%; Score 165; DB 1; Length 797;
Best Local Similarity 28.8%; Pred. No. 0.0033;
Matches 72; Conservative 37; Mismatches 89; Indels 52; Gaps 13;

Qy 22 LAL--DNSRNEGKLEALTDEFELEFLSKINGLTSISDL-----PKLKRLKE 69
| | | : | : | : | | | | | | | | | | | | | | | | | | | | |
Db 239 LALHPNDKASYSNILKFLTSNSHREHVTKVNVKAFMQSSLYKKIITKDKNPSPTKMLM 298
| | | : | : | : | | | | | | | | | | | | | | | | | | | | |
Qy 70 LRVSG-----GLEVLAE-----KCPNLTHLYLSONKINDLSTIEPL--KOLENLKSLDLFNC 119
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 299 ISILSRGIGMDLFCQSVLKAPLIDH-----KLSPVSEYEDFDEVELCISDDEVD 352
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 120 EVTN---LNDYGENVFKLL--QLTYLDSCYWDHKKAPYSDIEDHVEGLDDEEGEHE--- 172
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 353 EDGNLCVLDDSESVNSVALRQVLTVDK--QANEKEYKKIITKSDRDRDKDEYLEN 409
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 173 EYV--DEDAQVVEDEGE--EEEEEGEEDVSGDEDEEGYNDG-----EVGGE 218
| | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db 410 EYVNRDEEDEDGEDEKDEKDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDE 469
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 219 EDEELGEE 228
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 470 DDEDEGEDE 479
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 11
NFL_HUMAN STANDARD; PRT; 543 AA.
ID NFL_HUMAN
AC P07196; Q16154;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet L protein (68 kDa neurofilament protein)
DE (Neurofilament light polypeptide) (NF-L).
GN NEFL OR NFL OR NF68.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87214213; PubMed=3034332;
RA Julien J.-P., Grosveld F., Yazdankash K., Flavell D., Meijer D.,
RA Mushynski W.;
RT "The structure of a human neurofilament gene (NF-L): a unique exon-
RT intron organization in the intermediate filament gene family.";
RL Biochim. Biophys. Acta 909:10-20(1987).
RN [2]
RP SEQUENCE OF 1-10 FROM N.A.
RX MEDLINE=94235564; PubMed=8180132;
RA Pospelov V.A., Pospelova T.V., Julien J.-P.;
RT "AP-1 and Krox-24 transcription factors activate the neurofilament
RT light gene promoter in p19 embryonal carcinoma cells.";
RL Cell Growth Differ. 5:187-196(1994).
RN [3]
RP VARIANT CMT2E PRO-332.
RX MEDLINE=20307176; PubMed=10841809;
RA Mersisyanova I.V., Perepelov A.V., Polyakov A.V., Sitnikov V.F.,
RA Dadali E.L., Oparin R.B., Petrin A.N., Evgrafov O.V.;

```





|   |  |   |           |                               |
|---|--|---|-----------|-------------------------------|
| FT  | REPEAT   | 235   | 258       | LRR 4.                        |
| FT  | REPEAT   | 292   | 319       | LRR 5.                        |
| FT  | REPEAT   | 320   | 343       | LRR 6.                        |
| FT  | DOMAIN   | 359   | 397       | ASP/GLU-RICH (HIGHLY ACIDIC). |
| SQ  | SEQUENCE   | 587 AA;   | 63541 MW; | 3C18068AAC06B98F CRC64;       |
| Query Match   |  |   |           |                               |
| Best Local Similarity 28.8%; Score 160; DB 1; Length 587;       |  |   |           |                               |
| Matches 66; Conservative 30; Mismatches 47; Indels 86; Gaps 12; |  |   |           |                               |
| QY  | 25   | DNSRSNEGKLEALTDFFEEFLSKIN-----GGTSTISD-----LPKLIK-----LRKL    | 68        |                               |
| DB  | 245  | DNTFTKGAVALMAETFLKTLRQVEVFNFGDCLVRSGKVAIAADAIRGGLPKLIKELNLSFC | 303       |                               |
| QY  | 69   | ELRVSGGL---EVLARKCNLTHLYLSGNKIKIDLTIEPLKOLENLKSLDLFNCVNTNIN   | 125       |                               |
| DB  | 304  | EIKRDAALAAVAEMADKA-ELEKLDLNGNTLGE---EGCEQLQEVLE-----          | 346       |                               |
| QY  | 126  | DYGENVFKLLLOLTLYLDSYWDHKEAPYSIDIEDHVEGLDDEEGEHEEYEDDAQWVEDE   | 185       |                               |
| DB  | 347  | --GFNMAKVLASLS-----DDEDE-----EEEE                             | 367       |                               |
| QY  | 186  | EGEEEEEGEEDVSGDDEDEEGYNDGEVDEDEEELGEEERGQKRK                  | 234       |                               |
| DB  | 368  | EGEEEEEEAE-----EEDEE-----EEEEEEEEEPQQRQGEK                    | 406       |                               |
| RESULT 15   |  |   |           |                               |
| ASFL_YEAST  | ASFL_YEAST   | STANDARD;   | PRT;      | 279 AA.                       |
| AC  | P32447;  |   |           |                               |
| DT  | 01-OCT-1993 (Rel. 27, Created)   |   |           |                               |
| DT  | 01-OCT-1993 (Rel. 27, Last sequence update)  |   |           |                               |
| DT  | 16-OCT-2001 (Rel. 40, Last annotation update)  |   |           |                               |
| DE  | Anti-silencing protein 1.  |   |           |                               |
| GN  | ASF1 OR YJL115W OR J0755.  |   |           |                               |
| OS  | Saccharomyces cerevisiae (Baker's yeast).  |   |           |                               |
| OC  | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;   |   |           |                               |
| OC  | Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  |   |           |                               |
| OX  | NCBI_TaxID=4932;   |   |           |                               |
| RN  | [1]  |   |           |                               |
| RP  | SEQUENCE FROM N.A.   |   |           |                               |
| RC  | STRAIN=S288C / FY1679;   |   |           |                               |
| RC  | MEDLINE=97435476; PubMed=8948101;  |   |           |                               |
| RA  | Le S., Davis C., Konopka J.B., Sternglanz R.;  |   |           |                               |
| RT  | "Two new S-phase-specific genes from Saccharomyces cerevisiae.";   |   |           |                               |
| RL  | Yeast 13:1029-1042(1997).  |   |           |                               |
| RL  | [2]  |   |           |                               |
| RP  | SEQUENCE FROM N.A.   |   |           |                               |
| RC  | STRAIN=S288C / FY1679;   |   |           |                               |
| RC  | MEDLINE=97103775; PubMed=8948101;  |   |           |                               |
| RA  | Cziepluch C., Kordes E., Pujol A., Jauniaux J.-C.;   |   |           |                               |
| RT  | "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X   |   |           |                               |
| RT  | reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,  |   |           |                               |
| RT  | SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, GZF3, two tRNA genes,   |   |           |                               |
| RT  | three remnant delta elements and a Ty4 transposon.";   |   |           |                               |
| RL  | Yeast 12:1471-1474(1996).  |   |           |                               |
| CC  | -1- FUNCTION: DEREPRESSION OF SILENT MATING TYPE LOCI WHEN   |   |           |                               |
| CC  | OVEREXPRESSED.   |   |           |                               |
| CC  | -1- SIMILARITY: TO S.POMBE SPCC663.05C.  |   |           |                               |
| CC  | -----  |   |           |                               |
| CC  | This SWISS-PROT entry is copyright. It is produced through a collaboration   |   |           |                               |
| CC  | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |   |           |                               |
| CC  | the European Bioinformatics Institute. There are no restrictions on its  |   |           |                               |
| CC  | use by non-profit institutions as long as its content is in no way   |   |           |                               |
| CC  | modified and this statement is not removed. Usage by and for commercial  |   |           |                               |
| CC  | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |   |           |                               |
| CC  | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).  |   |           |                               |
| CC  | -----  |   |           |                               |
| DR  | EMBL; L07593; AAC37512.1; "  |   |           |                               |
| DR  | EMBL; 249390; CAA89410.1; "  |   |           |                               |
| DR  | PIR; S30766; S30766.   |   |           |                               |
| DR  | SGD; S0003651; ASFL.   |   |           |                               |
| FT  | DOMAIN   | 170   | 242       | ASP/GLU-RICH (HIGHLY ACIDIC). |



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2003, 14:53:13 ; Search time 88 Seconds  
(without alignments)

547.898 Million cell updates/sec

Title: US-09-591-500A-4

Perfect score: 1216

Sequence: 1 MEMGRIHSELNRAPSDVK.....VDGEDEELGEERQK 234

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Match | Length | DB | ID     | Description         |
|------------|-------|-------|--------|----|--------|---------------------|
| 1          | 1216  | 100.0 | 234    | 4  | O43423 | O43423 homo sapien  |
| 2          | 889   | 73.1  | 247    | 11 | O35381 | O35381 mus musculus |
| 3          | 885   | 72.8  | 247    | 11 | P97437 | P97437 mus musculus |
| 4          | 693.5 | 56.2  | 251    | 4  | O92688 | O92688 homo sapien  |
| 5          | 667   | 54.9  | 272    | 11 | O9EST6 | O9est6 rattus norv  |
| 6          | 655   | 53.9  | 272    | 11 | O9EST5 | O9est5 mus musculus |
| 7          | 643   | 52.9  | 200    | 11 | O9CXT4 | O9cxt4 mus musculus |
| 8          | 618   | 50.8  | 268    | 4  | O9BT0  | O9bt0 homo sapien   |
| 9          | 614   | 50.5  | 260    | 11 | P97822 | P97822 mus musculus |
| 10         | 610   | 50.2  | 266    | 4  | O8WWW9 | O8www9 homo sapien  |
| 11         | 609   | 50.1  | 260    | 11 | O9CZD2 | O9czd2 mus musculus |
| 12         | 490   | 40.3  | 131    | 4  | O95626 | O95626 homo sapien  |
| 13         | 402   | 33.1  | 261    | 5  | O9W895 | O9w895 drosophila   |
| 14         | 294.5 | 24.2  | 447    | 10 | O9SCQ7 | O9scq7 arabidopsis  |
| 15         | 243   | 20.0  | 229    | 5  | O62220 | O62220 caenorhabdi  |
| 16         | 230   | 18.9  | 225    | 5  | O01615 | O01615 caenorhabdi  |

|    |       |      |      |    |        |                    |
|----|-------|------|------|----|--------|--------------------|
| 17 | 176   | 14.5 | 409  | 10 | O9LN21 | O9ln21 arabidopsis |
| 18 | 174   | 14.3 | 567  | 4  | O9HD28 | O9hd28 homo sapien |
| 19 | 173.5 | 14.3 | 1152 | 5  | O8T216 | O8t216 dictyosteli |
| 20 | 171   | 14.1 | 1036 | 12 | O9DUM3 | O9dum3 kaposi's sa |
| 21 | 170.5 | 14.0 | 792  | 12 | O9DTL7 | O9ytl7 ateline her |
| 22 | 170   | 14.0 | 976  | 12 | O9DUN0 | O9dun0 kaposi's sa |
| 23 | 168   | 13.8 | 1089 | 12 | O40947 | O40947 kaposi's sa |
| 24 | 168   | 13.8 | 1129 | 12 | O9QR71 | O9qr71 kaposi's sa |
| 25 | 165   | 13.6 | 896  | 5  | O62348 | O62348 caenorhabdi |
| 26 | 165   | 13.6 | 906  | 5  | O45718 | O45718 caenorhabdi |
| 27 | 164   | 13.5 | 1019 | 3  | O94267 | O94267 schizosacch |
| 28 | 163.5 | 13.4 | 641  | 6  | O9SLM9 | O9slm9 macaca fasc |
| 29 | 163.5 | 13.4 | 934  | 6  | O9GMD3 | O9gmd3 bos taurus  |
| 30 | 163.5 | 13.4 | 1162 | 12 | O98148 | O98148 kaposi's sa |
| 31 | 163   | 13.3 | 3658 | 10 | O9W7K6 | O9w7k6 arabidopsis |
| 32 | 162   | 13.3 | 416  | 12 | O9YPA9 | O9ypa9 kaposi's sa |
| 33 | 161.5 | 13.3 | 386  | 4  | O8TCR7 | O8tc7 homo sapien  |
| 34 | 161   | 13.2 | 673  | 5  | O9U0N1 | O9u0n1 plasmodium  |
| 35 | 161   | 13.2 | 1003 | 12 | O9ILX9 | O9ilx9 kaposi's sa |
| 36 | 161   | 13.2 | 1127 | 3  | O9P571 | O9p571 neosporea   |
| 37 | 160   | 13.2 | 623  | 4  | O96J72 | O96j72 homo sapien |
| 38 | 160   | 13.2 | 798  | 13 | O90307 | O90307 carassius a |
| 39 | 159.5 | 13.1 | 1300 | 12 | O36421 | O36421 alceaphine  |
| 40 | 157.5 | 13.0 | 538  | 11 | O9ET15 | O9eti5 mus musculu |
| 41 | 157   | 12.9 | 495  | 12 | O918P0 | O918p0 ovine herpe |
| 42 | 157   | 12.9 | 1021 | 4  | O15451 | O15451 homo sapien |
| 43 | 157   | 12.9 | 1251 | 4  | O15450 | O15450 homo sapien |
| 44 | 156.5 | 12.9 | 791  | 13 | O9DGL1 | O9dgl1 fugu rubrip |
| 45 | 156   | 12.8 | 81   | 10 | O9LZR7 | O9l2r7 arabidopsis |

#### ALIGNMENTS

RESULT 1  
O43423 PRELIMINARY; PRT; 234 AA.  
AC O43423;  
DT 01-JUN-1998 (TRENBLrel. 06, Created)  
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE Pp32r1.  
GN Pp32r1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kadkol S.S., Brody J.R., Pevsner J., Bai J., Pasternack G.R.;  
RT "Modulation of oncogenic potential by alternative gene use in human  
prostate cancer.";  
RL Nat. Med. 0:0-0(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kochevar G.J., Brody J.R., Pasternack G.R.;  
RT "The Structure of a Gene Encoding pp32r1, a New Member of the pp32  
Family.";  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF008216; AAD12746.1;  
DR InterPro; IPR001611; LRR.  
DR Pfam; PF00560; LRR; 1.  
SQ SEQUENCE 234 AA; 26762 MW; 2515D9678EACC5C2 CRC64;

Query Match 100.0%; Score 1216; DB 4; Length 234;  
Best Local Similarity 100.0%; Pred. No. 4.1e-75;  
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEMGRIHSELNRAPSDVKELALDNRNECKLEALTDFFEEFLSKINGLTSIDL 60  
|||||  
Db 1 MEMGRIHSELNRAPSDVKELALDNRNECKLEALTDFFEEFLSKINGLTSIDL 60  
|||||  
QY 61 PKLKRKLELRVSGGLEVLAEKCPNLTLYLSGNKIKDLSTIEPLKQLKLENKSLDLFNC 120

```

Db      61  PKLKLRLVSGGLEVAEKCPNLTHLYLSGNKIKDLSTIEPLKQLENKLSLDLFNCE 120
QY      121  VTNLNDYGENVFKLLQLLTLDSCYWDHKEAPYSDIEDHVEGLDDEEGHEEYDQ 180
Db      121  VTNLNDYGENVFKLLQLLTLDSCYWDHKEAPYSDIEDHVEGLDDEEGHEEYDQ 180
QY      181  VVEDEGEDEEEDVSGDEDEEGYNDGEVDGEDEEELGEEERQKRK 234
Db      181  VVEDEGEDEEEDVSGDEDEEGYNDGEVDGEDEEELGEEERQKRK 234

RESULT 2
O35381  PRELIMINARY; PRT; 247 AA.
AC O35381;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cerebellar leucine rich acidic nuclear protein.
GN ANP32 OR LAMP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC TISSUE=CEREBELLUM;
RX MEDLINE=98013170; PubMed=9353121;
RA Matilla A., Koshiy B.T., Cummings C.J., Isobe T., Orr H.T.,
RA Zoghbi H.Y.;
RT "The cerebellar leucine-rich acidic nuclear protein interacts with
RT ataxin-1."
RL Nature 389:974-978(1997).
DR EMBL; AF022957; AAB91546.1; -.
DR MGD; MGI:108447; Anp32.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003603; LRRcap.
DR Pfam; PF00560; LRR; 2.
DR SMART; SM00446; LRRcap; 1.
KW Nuclear protein.
SQ SEQUENCE 247 AA; 28537 MW; 82EEDCF7ECC2918 CRC64;

Query Match 73.1%; Score 889; DB 11; Length 247;
Best Local Similarity 77.7%; Pred. No. 6.1e-53;
Matches 185; Conservative 16; Mismatches 31; Indels 6; Gaps 3;
QY 1 MEMGRRHSELNRAPSDVKELALDNRSGNEKLEALTDDEFEFLSKINGLTSISDL 60
Db 1 MEMDKRIYELNRNTPSDVKELVLDNCKSIEGKIEGLTDEFEFLSTINVLGTSISNL 60
QY 61 PKL-KLRKLEL---RVSGGLEVAEKCPNLTHLYLSGNKIKDLSTIEPLKQLENKLSLDL 116
Db 61 PKLKLRLVSGGLEVAEKCPNLTHLYLSGNKIKDLSTIEPLKQLENKLSLDL 120
QY 117 FNCVETNLNDYGENVFKLLQLLTLDSCYWDHKEAPYSDIEDHVEGLDDEEGHEEYD 176
Db 121 FNCVETNLNAYRENVEFKLLPQVMYLDGYDRDNKEAPSDSVEGYVE--DDDEDEDEEYD 178
QY 177 EDAQVVEDEGEDEEEDVSGDEDEEGYNDGEVDGEDEEELGEEERQKRK 234
Db 179 EYAQLVEDEEEDVSGDEDEEGYNDGEVDGEDEEELGEEERQKRK 236

RESULT 3
P97437  PRELIMINARY; PRT; 247 AA.
AC P97437;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Acidic nuclear phosphoprotein pp32.
GN ANP32.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RA Chen T.-H., Brody J.R., Romantsev F.E., Yu J.-G., Voneiff E.,
RA Kayler A.E., Kuhajda F.P., Pasternack G.R.;
RT "Structure of pp32, an Acidic Nuclear Protein Which Inhibits Oncogene-
RT Induced Formation of Transformed Foci.";
RL Mol. Biol. Cell 0:0-0(1996).
DR EMBL; U73478; AAB39707.1; -.
DR MGD; MGI:108447; Anp32.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003603; LRRcap.
DR Pfam; PF00560; LRR; 2.
DR SMART; SM00446; LRRcap; 1.
SQ SEQUENCE 247 AA; 28521 MW; 90PCDCE53CCC2918 CRC64;

Query Match 72.8%; Score 885; DB 11; Length 247;
Best Local Similarity 77.7%; Pred. No. 1.1e-52;
Matches 185; Conservative 15; Mismatches 32; Indels 6; Gaps 3;
QY 1 MEMGRRHSELNRAPSDVKELALDNRSGNEKLEALTDDEFEFLSKINGLTSISDL 60
Db 1 MEMDKRIYELNRNTPSDVKELVLDNCKSIEGKIEGLTDEFEFLSTINVLGTSISNL 60
QY 61 PKL-KLRKLEL---RVSGGLEVAEKCPNLTHLYLSGNKIKDLSTIEPLKQLENKLSLDL 116
Db 61 PKLKLRLVSGGLEVAEKCPNLTHLYLSGNKIKDLSTIEPLKQLENKLSLDL 120
QY 117 FNCVETNLNDYGENVFKLLQLLTLDSCYWDHKEAPYSDIEDHVEGLDDEEGHEEYD 176
Db 121 FNCVETNLNAYRENVEFKLLPQVMYLDGYDRDNKEAPSDSVEGYVE--DDDEDEDEEYD 178
QY 177 EDAQVVEDEGEDEEEDVSGDEDEEGYNDGEVDGEDEEELGEEERQKRK 234
Db 179 EYAQLVEDEEEDVSGDEDEEGYNDGEVDGEDEEELGEEERQKRK 236

RESULT 4
Q92688  PRELIMINARY; PRT; 251 AA.
AC Q92688; O00655; P78459; P78458;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE PHAPI2A protein (APRIL) (PHAPI2B protein) (Acidic protein rich in
DE leucines) (Similar to acidic protein rich IN LEUCINES).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RA Vaesen M., Barnikol-Watanabe S., Kratzin H.D., Hilschmann N.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN [2]
SEQUENCE FROM N.A.
RA Zhu L., Henning D., Valdez B.C.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RN [3]
SEQUENCE OF 3-251 FROM N.A.
RC TISSUE-PANCREAS;
RA Mencinger M., Panagopoulos I., Contreras J.A., Mitelman F., Aman P.;
RL Genomics 0:0-0(0).
RN [4]
RN [4]
SEQUENCE FROM N.A.
RC TISSUE-MUSCLE;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RN [5]
SEQUENCE FROM N.A.

```

```
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYONIC TESTIS CARCINOMA;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y07569; CAA68855.1; -
DR EMBL; U70439; AAB37579.1; -
DR EMBL; Y07969; CAA69265.1; -
DR EMBL; Y07570; CAA68856.1; -
DR EMBL; BC000476; AAH00476.1; -
DR EMBL; BC013003; AAH13003.1; -
DR EMBL; BC019658; AAH19658.1; -
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003603; LRRcap.
DR Pfam; PF00560; LRR; 2.
DR SMART; SM00446; LRRcap; 1.
SQ SEQUENCE 251 AA; 28787 MW; 93A1AADF8EDE7D53 CRC64;

Query Match 56.2%; Score 683.5; DB 4; Length 251;
Best Local Similarity 61.0%; Pred. No. 5.2e-39;
Matches 152; Conservative 34; Mismatches 40; Indels 23; Gaps 8;

QY 1 MEMGRIHSELNRNAPSVDVKELALDNRSGNEKLEALTDEFELEFLSKINGLTSIDL 60
DQ 1 MEMGRIHSELNRNAPSVDVKELALDNRSGNEKLEALTDEFELEFLSKINGLTSIDL 60
DQ 1 MDKMRRIHLELRNRTPAAVRELVDNCKSDCKGKIEGLTAEFVNLFLSLINVLVSMDL 60
QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTHLYLSGNKIKDLSLTIPLKLENKLSIDL 116
DQ 61 PKLPLKLELSENRIFGGLDMLAEKLPNLTHLNSGNKIKDLSLTIPLKLECLKLSIDL 120
QY 117 FNCVNTNLDYGENVFKLLQLTYLDSCYWDHKEAPYSIDIEDHVGLEDDEEGEH-EEY 175
DQ 121 FNCVNTNLDYGENVFKLLQLTYLDSCYWDHKEAPYSIDIEDHVGLEDDEEGEH-EEY 175
DQ 121 FNCVNTNLDYGENVFKLLQLTYLDSCYWDHKEAPYSIDIEDHVGLEDDEEGEH-EEY 175
QY 176 DEDAQVVEDEEGEE---EEEGEEEDVSG-----GDEDEEGYNDGVEDEEGEELG 225
DQ 179 DED-----DEEGEEEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 232
QY 226 EERGOKRK 234
DQ 233 EGGKGEKK 241

RESULT 5
Q9EST6 PRELIMINARY; PRT; 272 AA.
AC Q9EST6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Proliferation related acidic leucine rich protein PAL31.
OS Rattus norvegicus (Rat)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR-IMMICH; TISSUE=BRIN;
RX MEDLINE=21092721; PubMed=11162633;
RA Sun W., Hattori N., Mutai H., Toyoshima Y., Kimura H., Tanaka S.,
RA Shiota K.;
RT "PAL31, a nuclear protein required for progression to the S phase.";
RL Biochem. Biophys. Res. Commun. 280:1048-1054(2001).
DR EMBL; AB025581; BAB12435.1; -
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003603; LRRcap.
DR Pfam; PF00560; LRR; 2.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00446; LRRcap; 1.
SQ SEQUENCE 272 AA; 31060 MW; 556AC9B169E9D996 CRC64;

Query Match 53.9%; Score 655; DB 11; Length 272;
Best Local Similarity 54.7%; Pred. No. 4.8e-37;
Matches 145; Conservative 40; Mismatches 46; Indels 34; Gaps 7;

QY 1 MEMGRIHSELNRNAPSVDVKELALDNRSGNEKLEALTDEFELEFLSKINGLTSIDL 60
DQ 1 MDKMRRIHLELRNRTPAAVRELVDNCKKAMDGKIEGLTDEFVNLFLSLINVLFSVSDL 60
QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTHLYLSGNKIKDLSLTIPLKLENKLSIDL 116
DQ 61 PKLPLKLELSENRIFGGLDMLAEKLPNLTHLNSGNKIKDLSLTIPLKLECLKLSIDL 120
QY 117 FNCVNTNLDYGENVFKLLQLTYLDSCYWDHKEAPYSIDIEDHVGLEDDEEGEH-EEY 165
DQ 121 FNCVNTNLDYGENVFKLLQLTYLDSCYWDHKEAPYSIDIEDHVGLEDDEEGEH-EEY 180
QY 166 EEEGE---HEEYDEDAQVVEDEEGEEEGEEEGEEEGEEEGEEEGEEEGEEEGEE 215
DQ 181 EEEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 237
QY 216 DGEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 234
DQ 238 DEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 262

RESULT 6
Q9EST5 PRELIMINARY; PRT; 272 AA.
AC Q9EST5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Proliferation related acidic leucine rich protein PAL31 (Similar to
DE acidic protein rich in leucines).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALBCA JCL; TISSUE=BRIN;
RA Mutai H., Toyoshima Y., Sun W., Takigawa H., Tanaka S., Shiota K.;
RT "Expression of a Novel Gene, Proliferation Related Acidic Leucine Rich
RT Protein (PAL31), in the Developing Brain.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB025582; BAB12436.1; -
DR EMBL; BC005628; AAH05628.1; -
DR EMBL; BC003489; AAH03489.1; -
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003603; LRRcap.
DR Pfam; PF00560; LRR; 2.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00446; LRRcap; 1.
SQ SEQUENCE 272 AA; 31078 MW; BC22DC591AA6BC15 CRC64;

Query Match 53.9%; Score 655; DB 11; Length 272;
Best Local Similarity 54.7%; Pred. No. 4.8e-37;
Matches 145; Conservative 40; Mismatches 46; Indels 34; Gaps 7;

QY 1 MEMGRIHSELNRNAPSVDVKELALDNRSGNEKLEALTDEFELEFLSKINGLTSIDL 60
DQ 1 MDKMRRIHLELRNRTPAAVRELVDNCKKAMDGKIEGLTDEFVNLFLSLINVLFSVSDL 60
```

QY 61 PKL-KLRKLEL---RVSGGLEVLAEKCPNLTHLYSGNKIKDLSTIEPLKOLENKLKSLDL 116  
DB 61 PKLPLKLELSENIFGLRLAEELSLTHLNLGNLKDISTIEPLKRLDLCKLSLDL 120  
QY 117 FNCVETNLNDYGENVFKLLQLTYLDSYWDHKEAPYSDIEHDVHVEGLDDEREGEHEEYDEDA 165  
DB 121 FCEVTNRSYDRETVERLLPQLSYLDGYDRDQAPDSVBEVSDVEERAPDSGVDGVOK 180  
QY 166 EEEGE---HEEYDEDAQVDEDEGEHEEYDEEYSGGDEED-----DHVEGLDD 215  
DB 181 EEEDEEHEEEDG---EEDEDEDEDEDEVEGEDEDEVEGSEEEHFGHGEV 237  
QY 216 DCEDEEELGEE-----RGOKRK 234  
DB 238 DEDEDEDEDEDEEESGKGK 262

## RESULT 7

Q9CXT4 PRELIMINARY; PRT; 200 AA.  
AC Q9CXT4; 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DE 13 days embryo head cDNA, RIKEN full-length enriched library,  
DE clone:3110006M19, full insert sequence.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gofjoberi T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., Kigle B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,  
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Offitt T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Bojelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Cariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
PL Nature 409:685-690(2001).  
DR EMBL; AK014013; BAB29114.1; -  
DR InterPro; IPR001611; LRR.  
DR InterPro; IPR003603; LRRcap.  
DR Pfam; PF00560; LRR; 2.  
DR PRINTS; PR00019; LEURICHRPT.  
DR SMART; SM00446; LRRcap; 1.  
SQ SEQUENCE 200 AA; 23161 MW; 6284B81564F108FF CRC64;

Query Match 52.9%; Score 643; DB 11; Length 200;  
Best Local Similarity 72.2%; Pred. No. 2.2e-36;  
Matches 140; Conservative 15; Mismatches 33; Indels 6; Gaps 3;

QY 4 GRRHSELNRAPSDVKEALDNRSGNEKLEALDTEFELEFLSKINGLTSISLPLK 63  
DB 7 GRRWAASWIKQKQKGVGLVDNCKSTEGKIEGLTDEFELEFLSKINGLTSISLPLK 66  
QY 64 -KLRKLEL---RVSGGLEVLAEKCPNLTHLYSGNKIKDLSTIEPLKOLENKLKSLDLFNC 119  
DB 67 NKLKLELSENISGDLVLAEKCPNLKHLNLSGNKIKDLSTIEPLKLELKLKSLDLFNC 126

QY 120 EVTNLDYGENVFKLLQLTYLDSYWDHKEAPYSDIEHDVHVEGLDDEREGEHEEYDEDA 179  
DB 127 EVTNLNRAYENVFKLLPQWYLDGYDRNKEAPSDVEGYVE--DDDEDEDEEYDEYA 184  
QY 180 QVVEDEGEHEE 193  
DB 185 QLVDEDEEDEE 198

## RESULT 8

Q9BTTO PRELIMINARY; PRT; 268 AA.  
AC Q9BTTO; 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Similar to RIKEN cDNA 2810018A15 gene.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=CERVIX;  
RA Strausberg R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC003380; AA03380.1; -  
DR InterPro; IPR001611; LRR.  
DR InterPro; IPR003603; LRRcap.  
DR Pfam; PF00560; LRR; 2.  
DR SMART; SM00446; LRRcap; 1.  
SQ SEQUENCE 268 AA; 30692 MW; 99D74AF4B59BF971 CRC64;

Query Match 50.8%; Score 618; DB 4; Length 268;  
Best Local Similarity 54.2%; Pred. No. 1.5e-34;  
Matches 141; Conservative 37; Mismatches 52; Indels 30; Gaps 7;

QY 1 MEMGRRHSELNRAPSDVKEALDNRSGNEKLEALDTEFELEFLSKINGLTSISDL 60  
DB 1 MEMKKINLELNRSPPEETELVDNLCVNGEGLDFTKELEFLSMANVELSSIALR 60  
QY 61 PKL-KLRKLEL---VSGGLEVLAEKCPNLTHLYSGNKIKDLSTIEPLKOLENKLKSLDL 116  
DB 61 PSLNKLKLELSDNIISGGLVLAEKCPNLTHLYSGNKIKDLSTVEALQNLKLSLDL 120  
QY 117 FNCVETNLNDYGENVFKLLQLTYLDSYWDHKEAPYSDIEHDVHVEGLDDEREGEHEE--- 172  
DB 121 FNCETNLNDYRESITFELLOQITLYLDGFDQEDNEAPDSEEDDEDEDEDEEENAGP 180  
QY 173 -EEYDEDAQVVEDEGEHEEYDEEYSG-----GDEE-----DEEYNDGE 214  
DB 181 PEGYEE-----EEEEDEDEDEDEAGSELGEEVGLSYLMKEEQDEDDDDYV 236  
QY 215 VDGEDEEELGEEERQK 234  
DB 237 BEGEDEEELGEEERQK 256

## RESULT 9

P97822 PRELIMINARY; PRT; 260 AA.  
AC P97822; 01-MAY-1997 (Tremblrel. 03, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Cerebellar postnatal development protein-1 (LAMP-like protein) (RIKEN  
DE cDNA 2810018A15 gene).  
GN CLA3 OR 2810018A15RIK OR CPD1 OR MLANP-L.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

```

RN RP SEQUENCE FROM N.A.
RC STRAIN-C57BL6J; TISSUE-CEREBELLUM;
RA Radrizzani M., Vila-Ortiz G.J., Caferata E.G.A.,
RA Gonzalez Guerrero A.M., Di Tella M.C., Pivetta O.H., Carminatti H.,
RA Itoyaga-Vargas V.P., Santa-Coloma T.A.;
RT "Differential Expression of CpD1 During Postnatal Development in Mouse
RT Cerebellum.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN RP
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN-C57BL6; TISSUE-BRAIN;
RA Matsubae M.;
RT "Characterization of the nuclear transport a novel leucine-rich acidic
RT nuclear protein-like protein.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN RP
[3]
RN RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89345; ABA9462.2; -
DR EMBL; AB037685; BAB03507.1; -
DR EMBL; BC005690; AAH05690.1; -
DR MGD; MGI:1913721; Cla3.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003603; LRRcap.
DR Pfam; PF00560; LRR; 2
DR SMART; SM00446; LRRcap; 1.
SQ SEQUENCE 260 AA; 29622 MW; 7F94E46D72A04780 CRC64;

Query Match 50.5%; Score 614; DB 11; Length 260;
Best Local Similarity 53.3%; Pred. No. 2.7e-34;
Matches 139; Conservative 35; Mismatches 47; Indels 40; Gaps 8;

QY 1 MEMGRIHSELNRAPSDVKELALDNRSGKLEALTFEFLSKINGLTSISDL 60
Db 1 MEMKKINLELRNRPSPVEVTELVLDNCLVNGEIEGLNDFKEFLSMANVELSLARL 60
1 MEMKKINLELRNRPSPVEVTELVLDNCLVNGEIEGLNDFKEFLSMANVELSLARL 60
QY 61 PKL-KLRKLELR---VSGGLEVLAKCPNLTHLYSGNKIKDLSLTIPLKQLENKLSLDL 116
Db 61 PSLNKLRLKLELSDNIISGGLVLAECPNLTLYNLGSKNKIKDLSLTIPLKQLENKLSLDL 120
117 FNCVETNLNDYGENVFKLLQLTLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYD 176
Db 121 FNCVETNLNDYRESIFELQIITYLDGFDQEDNEAPDSEEDDEGDEDEE-EEENRAG 179
177 EDAQVVEDEGEDEEEDVSG-----GDEE-----DEEGYNDGEVDEE 219
Db 180 PPEGYEEDDEEEDDEDEDEAGSELGEEVGLSYLMKEEIQDEEDDDDDYVEEGEE 239
220 DEELGEEERQKRR 234
Db 240 EEEEEEGGLGEKRR 254

RESULT 11
Q9CZD2 PRELIMINARY; PRT; 260 AA.
ID Q9CZD2;
AC Q9CZD2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 2810018A1Srik protein.
GN CLA3 OR 2810018A1SRK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Bozell D., Bojunga N., Carninci P., de Bonaudo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK012759; BAB28449.1; -
DR MGD; MGI:1913721; Cla3.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003603; LRRcap.
DR Pfam; PF00560; LRR; 2.
DR SMART; SM00446; LRRcap; 1.

```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN-C57BL6J; TISSUE-CEREBELLUM;
RA Radrizzani M., Vila-Ortiz G.J., Caferata E.G.A.,
RA Gonzalez Guerrero A.M., Di Tella M.C., Pivetta O.H., Carminatti H.,
RA Itoyaga-Vargas V.P., Santa-Coloma T.A.;
RT "Differential Expression of CpD1 During Postnatal Development in Mouse
RT Cerebellum.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN RP
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN-C57BL6; TISSUE-BRAIN;
RA Matsubae M.;
RT "Characterization of the nuclear transport a novel leucine-rich acidic
RT nuclear protein-like protein.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN RP
[3]
RN RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89345; ABA9462.2; -
DR EMBL; AB037685; BAB03507.1; -
DR EMBL; BC005690; AAH05690.1; -
DR MGD; MGI:1913721; Cla3.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003603; LRRcap.
DR Pfam; PF00560; LRR; 2
DR SMART; SM00446; LRRcap; 1.
SQ SEQUENCE 260 AA; 29622 MW; 7F94E46D72A04780 CRC64;

Query Match 50.5%; Score 614; DB 11; Length 260;
Best Local Similarity 53.3%; Pred. No. 2.7e-34;
Matches 139; Conservative 35; Mismatches 47; Indels 40; Gaps 8;

QY 1 MEMGRIHSELNRAPSDVKELALDNRSGKLEALTFEFLSKINGLTSISDL 60
Db 1 MEMKKINLELRNRPSPVEVTELVLDNCLVNGEIEGLNDFKEFLSMANVELSLARL 60
1 MEMKKINLELRNRPSPVEVTELVLDNCLVNGEIEGLNDFKEFLSMANVELSLARL 60
QY 61 PKL-KLRKLELR---VSGGLEVLAKCPNLTHLYSGNKIKDLSLTIPLKQLENKLSLDL 116
Db 61 PSLNKLRLKLELSDNIISGGLVLAECPNLTLYNLGSKNKIKDLSLTIPLKQLENKLSLDL 120
117 FNCVETNLNDYGENVFKLLQLTLTYLDSCYWDHKEAPYSDIEDHVEGLDDEEGEHEEYD 176
Db 121 FNCVETNLNDYRESIFELQIITYLDGFDQEDNEAPDSEED-----DDEGDEED-D 174
177 EDAQVVEDEGER---EEEGEEDVSG-----DEEDGYNDG 213
Db 175 ED----EDEAGPPGEEEDDEDEAGSEVGEVGLSYLMKDEIQDEEDDDYVD- 229
214 EVDGEDEELGEEERQKRR 234
Db 230 --EGEEEEEEGLGEKRR 248

RESULT 10
Q8WWN9 PRELIMINARY; PRT; 266 AA.
ID Q8WWN9;
AC Q8WWN9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Leucine-rich acidic protein-like protein.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mao Y., Xie Y., Jiang M.;
RT "Cloning and Characterization of a Novel Human LANP-L Gene.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY057381; AAL25814.1; -

```

```
DR SMART; SM00446; LRRCP; I.
SO SEQUENCE 261 AA: 29176 MW: 18BC42BDDDAED2EE CRC64;
```

```

Query Match      33.1%; Score 402; DB 5; Length 261;
Best Local Similarity 40.2%; Pred. No. 6.3e-20;
Matches 106; Conservative 38; Mismatches 82; Indels 38; Gaps 10;

QY 3 MGRRIHSELRNAPSVDVKELALDNRSRNECKLALTDPEFEELFKINGGLTISDLPK 62
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MEKKRIELERRARKVNOITELNDNCNST--SIVGLTDEYVALSGLSLINVLGTLTKGFKK 58
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



